OEI-10014

ROPES & GRAY
ONE FRANKLIN SQUARE
1301 K STREET, NW
SUITE 800 EAST
WASHINGTON, DC 20005-3333
PHONE: (202) 626-3900
FAX: (202) 626-3961
DIRECT DIAL: (202) 626-3905

June 14, 2002

Docket ID No. OEI-10014
U.S. Environmental Protection Agency
Northeast Mall (Room B607)
401 M Street SW
Washington, DC 20460

Re: Guidelines for Ensuring and Maximizing the Quality, Objectivity, Utility and Integrity of Information Disseminated by the Environmental Protection Agency

Dear To Whom It May Concern:

On behalf of the Coalition for Effective Environmental Information, I am submitting these comments on the Agency's "Data Quality Guidelines" that were proposed for comment on April 30, 2002 (67 Fed. Reg. 21234). Thank your for your attention to this matter.

Sincerely,

Mark A. Greenwood

Enclosure

7125038

Before the United States
Environmental Protection Agency

Draft Guidelines for Ensuring and Maximizing the Quality, Objectivity, Utility and Integrity of Information Disseminated by the Environmental Protection Agency

Docket ID No. OEI-10014

Comments of the Coalition for Effective Environmental Information

---

On April 30, 2002, the U.S. Environmental Protection Agency (EPA) issued a Federal Register Notice (67 Fed. Reg. 21234) requesting comment on its draft Guidelines for Ensuring and Maximizing the Quality, Objectivity, Utility and Integrity of Information Disseminated by the Environmental Protection Agency (EPA Guidelines or Guidelines). These EPA Guidelines were issued pursuant to statutory requirements under Section 515 of the Treasury and General Government Appropriations Act for Fiscal Year 2001 (Data Quality Act of 2000, Pub. L. No. 106-554, 114 Stat. 2763, H.R. 5658) and the Guidelines for Ensuring and Maximizing the Quality, Objectivity, Utility and Integrity of Information Disseminated by Federal Agencies (OMB Guidelines) issued by the Office of Management and Budget (OMB) on February 22, 2002 (67 Fed. Reg. 8452).[1]

The Coalition for Effective Environmental Information (CEEI) appreciates the opportunity to comment on the EPA Guidelines. CEEI is a group of major companies and business organizations, representing a wide range of industry sectors, that share a common interest in improving how the government collects, manages, uses and disseminates environmental information.[2] CEEI supports public policies that encourage data quality, governmental accountability, efficient data collection, alignment of data with strategic goals and consistent management of environmental information resources.

---

[1] The OMB Guidelines issued on February 22, 2002 were a republication, with minor editing, of final Guidelines that OMB issued on January 3, 2002 (67 Fed. Reg. 369). These Guidelines were originally proposed for comment on June 28, 2001 (66 Fed. Reg. 34489) and finalized on an interim basis on September 28, 2001 (66 Fed. Reg. 49718). OMB has also issued further guidance on the Guidelines in a memorandum to the President's Management Council, dated June 10, 2002 (OMB Memorandum).

[2] CEEI includes representatives from the aerospace, chemical, energy, automobile, pharmaceutical, forest products, petroleum, electronics and consumer products industries.

CEEI sees these Guidelines as an opportunity to institutionalize good practices for assembling and presenting information that guides public and private decisionmaking. Many of these good practices have been publicly endorsed by EPA over the last several years. The EPA Guidelines can foster implementation of good practices for data quality in several ways:

- Define the roles and responsibilities of key governmental actors in assuring data quality;
- Establish standard practices and an overall management system that facilitate data quality in EPA information disseminations;
- Establish lines of communication and dispute resolution mechanisms; and
- Provide a working definition of "data quality".[3]

Unfortunately, this version of the EPA Guidelines does not match our expectations for what the Agency could have proposed. We believe that this proposal represents a missed opportunity to provide a real implementation plan for achieving the data quality objectives that EPA has embraced. EPA has not addressed several key issues about data quality. For the issues that EPA has addressed, the Guidelines do not always provide effective approaches.

We are concerned that this version of the Guidelines has not adequately focused on what is needed for *implementation* of data quality objectives. EPA seems primarily oriented toward preserving "business as usual" in the managing of information by Headquarters program offices, Regional offices and state agencies. CEEI believes that the EPA Guidelines can be improved substantially and offers the following recommendations.

Role of the Guidelines

The draft EPA Guidelines do not clearly articulate their role and effect in the management of the Agency. Instead, the Guidelines primarily emphasize what they are not. In Section 1.1 of the Guidelines, EPA offers the following characterization in explaining the purpose of the Guidelines:

> The guidelines are not a regulation. They are not legally enforceable and do not create any legal rights or impose any legally binding requirements or obligations on EPA or the public. Nothing in these guidelines affects any otherwise available judicial review of EPA action. The guidelines

[3] CEEI believes that "data quality" is a broad concept that should encompass a range of attributes. We have discussed the elements of data quality in several contexts, including the attached paper entitled "The Data Quality Challenge" (December 2000) as well as CEEI's comments on the OMB Guidelines, which we have also attached.

> may not apply to a particular situation based on the circumstances, and EPA retains discretion to adopt approaches on a case-by-case basis that differ from the guidelines, where appropriate. Any decisions regarding a particular case, matter or action will be made based on applicable statutes, regulations and requirements.

This opening statement of the Guidelines, articulating the purpose for all the measures outlined in the rest of the document, sends the wrong signal about the practical importance of the Guidelines. This statement can be read as a signal to EPA employees that the Guidelines are not to be taken seriously. Certainly members of the public who had hoped that the Guidelines would be a touchstone for how EPA would manage data quality concerns can have no confidence that the Guidelines truly represent EPA policy.

While CEEI can understand EPA's need to retain some flexibility in implementation of its data quality program, the language outlined above would essentially allow the Agency to abandon the Guidelines whenever EPA thought that ignoring the Guidelines was "appropriate." Section 515 contemplates that OMB would issue Guidelines addressing all aspects of data quality and providing an administrative mechanism for the correction of data. The OMB Guidelines, in turn, define specific responsibilities that agencies must discharge. Neither the statute nor the OMB Guidelines authorize agencies to establish a general waiver of its Guidelines whenever such a waiver seems appropriate to the agency.[4] EPA should remove the open-ended waiver language from Section 1.1 of the Guidelines.[5]

Building Blocks for a Successful Data Quality Program

Both the preamble to the Guidelines and the Guidelines themselves emphasize four existing policies that constitute the foundation for the Agency's data quality program. These policies include the following:

- Agency-wide Quality System (EPA Order 5360.1);
- Peer Review Policy;
- Action Development Process; and

---

[4] Under the Data Quality Act, OMB is to issue its guidelines under 44 U.S.C. §3504(d)(1) and §3516 of the Paperwork Reduction Act (PRA). That law further provides, at 44 U.S.C. §3506(a)(1)(B), that the head of each federal agency must comply with the requirements and "related policies" issued by OMB under the PRA. Thus EPA does not have the option to waive itself out of the requirements of the OMB Guidelines.

[5] Section V. of the OMB Memorandum makes this point most clearly, stating, "Regardless of what kinds of litigation-oriented disclaimers the agencies may include, agency guidelines should not suggest that agencies are free to disregard their own guidelines."

- Integrated Error Correction Process.

These policies are at various stages of development and address differing aspects of the general criteria for a data quality program. They do not inherently provide an integrated policy, and certainly do not constitute a management system, to assure the quality of EPA information disseminations. In addition, some of these policies are old Agency policies that have not been fully implemented over time. CEEI also notes that the list does not include some promising new efforts in the Agency to establish a systematic approach for developing good information products.

CEEI urges EPA to avoid relying on a patchwork of existing policies, particularly those that have not been fully implemented, as its strategy for addressing data quality. EPA needs to evaluate existing policies, identify the areas where new policies are needed and design a more strategic approach to the data quality issue. The following comments address the strengths and weaknesses of existing policies:

1. Quality System – For many years EPA has had policies in place that are intended to assure that data collected by the Agency meet certain quality assurance and quality control standards. These policies, set forth in EPA Order 5360, are primarily focused on initial data collection as part of scientific studies, addressing issues of sampling, analytical methods and documentation of results. On its face EPA Order 5360 defines the roles and responsibilities of various Agency personnel, outlines specific tasks and specifies particular outputs (e.g., Quality Management Plan, Quality Assurance Project Plans.) Yet this Order is unlikely to have substantial effect on overall data quality in EPA programs.

While Order 5360 has existed for some time, EPA personnel have not always complied with its terms. Several General Accounting Office (GAO) reports have noted this pattern of uneven implementation of the Order. Agency personnel have not, for example, required statistically valid sampling plans in the hazardous waste or water programs.[6] In addition, a systematic review of the EPA water programs determined that EPA was not implementing Order 5360 in many of the core elements of those programs.[7]

---

[6] GAO, Environmental Enforcement: EPA Cannot Ensure the Accuracy of Self-Reported Compliance Monitoring Data, at 26-30 (March 1993).

[7] GAO, Water Pollution: Poor Quality Assurance and Limited Pollutant Coverage Undermine EPA's Control of Toxic Substances, at 33-45 (1994). This report evaluated the seven core elements of the water program's approach to toxic substances. The report concluded that, while EPA had good quality assurance in two of the seven areas, in the other five areas EPA complied with only 58% of the 40 key criteria for quality assurance.

Besides the general implementation problems that have historically occurred in EPA's quality assurance program, other systemic factors suggest that Order 5360 can have only limited impact on the data quality issues of the Agency. Most data in EPA files were collected by other parties, including states, private companies, independent laboratories and universities. These parties have defined their own quality assurance measures commensurate with their own needs and purposes. Use of the data by EPA may be only one of several purposes of the data.

Thus, EPA is seldom in a position to dictate how particular data will be collected. The more typical challenge for EPA is to consider how particular data sets should be used and disseminated by the Agency for the "secondary uses" that emerge from the changing agendas of EPA programs. As an example, under Section 305 (b) of the Clean Water Act EPA collects water quality monitoring and assessment data from the states. EPA assembles these data into the National Water Quality Inventory report (commonly referred to as the "305(b) Report".) The states utilize such different methodologies in their assessments that the data are not comparable from state to state.[8] GAO and some states have criticized EPA's inclination to use the 305(b) Report as a basis for national public databases, such as the Index of Watershed Indicators, in light of the differing state methodologies for collecting data.[9] EPA Order 5360 has not played a significant role in this matter or other major data quality controversies of the last several years.

2. *Peer Review Policy* – EPA has also had a peer review policy for several years. The policy provides considerable discretion for EPA program offices to determine when peer review is needed and the extent of peer review that should occur. Under the EPA policy, a "peer review" can be everything from an open public meeting in which peer reviewers hear testimony from interested parties and file a formal report to a brief review by a few unidentified contractors and EPA staff.

As a result, the mere existence of the pre-existing Peer Review Policy does not provide an assurance that the objectives of the OMB Guidelines will be met. If EPA plans to rely on the Peer Review Policy as a cornerstone of its approach to data quality, it should consider how the Policy might be modified to align it with the Guidelines.

Under the OMB Guidelines, peer review is a component of the "objectivity" standard. Specifically, the OMB Guidelines create a rebuttable presumption that data meets the "objectivity" standard when the data or analytical results have been subject to "formal, independent, external peer review."[10] The OMB Guidelines further define the "general criteria for competent and credible peer review" that must be met to qualify for the rebuttable presumption in the Guidelines.

---

[8] EPA's reporting guidelines under Section 305(b) are not mandatory on the states.

[9] GAO, Water Quality: Key EPA and State Decisions Limited by Inconsistent and Incomplete Data, at 36 (March 2000).

[10] OMB Guidelines, at V.3.b.i.

EPA should be revising its peer review policies to clarify when the rebuttable presumption applies. Besides the general disclosure requirements specified in the OMB Guidelines, EPA should also address when a peer review is sufficiently "formal" and "independent" to qualify for the rebuttable presumption. For example, it would not appear that a peer review conducted by Agency personnel or by unidentified contractors would meet the "formal and independent" standard.

EPA would also need to address how a peer review relates to the intended purposes of an information dissemination. For example, a peer review may occur when a particular data-set is used for its original purpose (e.g., whether a chemical meets the standard for "listing" within a particular program.) Many years later EPA may want to use that same data-set to draw a broader conclusion about a class of chemicals. The original peer review may not have addressed some of the key scientific questions relevant to the later "secondary use" of the data. In such a situation, the rebuttable presumption should no longer be applicable.

Thus, while EPA's Peer Review Policy is certainly relevant to the Guidelines, further work is needed before it can be a true cornerstone for EPA's data quality program.

3. Action Development Process – In every Administration EPA has had an internal process to initiate major rules and policies. As part of that process Agency managers look at the question of what data should be developed to support the proposed action. While this process has gone under several names, in this Administration it is called the "Action Development Process." This process is unlikely to be a central component of EPA's data quality program for several reasons.

First, each new Administration tends to rework this process, adapting the process to its particular needs and priorities. As a result, the process has not had continuity over time. While this Administration may focus on data quality issues during the process, the next Administration may articulate a different set of priorities. Second, it is unclear who makes final determinations about the adequacy of the plans for data development in the described Action Development Process. It appears that high quality data is expected to emerge from negotiations among program offices, regional offices, states and cross-office groups (e.g., Regulatory Policy Council, Science Policy Council.) This restatement of the status quo is unlikely to change the uneven quality in EPA's information dissemination activity.

Third, the Action Development Process is not a process that is open to the public. It is a process internal to EPA that does not offer systematic opportunities for interested stakeholders to participate. As a result, EPA will not always know what will be the key data quality issues of interest to stakeholders until the rulemaking process is far along.

4. Integrated Error Correction Process – Over the last two years EPA has established and expanded its efforts to provide for correction of errors in Agency

databases. This process addresses simple questions of data accuracy, rather than more complex questions of data presentation or characterization.

While this new process needs further refinement, CEEI believes that it can provide a useful building block for an effective data quality program. Under the Integrated Error Correction Process, EPA receives requests to modify information posted on its Website. Once an error correction request is made, the Office of Environmental Information (OEI) conveys the request to the appropriate "data owner" and tracks the process for resolution of the request.[11] CEEI believes that EPA has taken a reasonable step in developing this process and urges the Agency to refine its operation.

Industry experience with the Integrated Error Correction Process has generally been favorable. At the same time there have been situations where it takes an inordinate amount of time to resolve fairly simple corrections of inaccuracies.[12] While CEEI has not attempted to undertake a systematic assessment of the Process, anecdotal information suggests that the responsiveness of the system is highly dependent on the attitude and engagement of the "data owner". Where the data owner has been responsive, as in the case of data elements managed by OEI, the correction requests can often be addressed in a few days. In other situations, correction requests can languish for months if the data owner does not see the correction as a priority.

In developing its data quality program, EPA should give particular attention to how it can improve the response time of the Integrated Error Correction Process. Based on its experience to date, EPA should be able to establish reasonable response times that can be specified in Agency policy, including the Guidelines. To the extent that responsiveness from third parties, such as the states, has been a particular obstacle, EPA should consider making timely responses to correction requests an explicit condition of the Agency state grant programs.

5. *Information Product Development Process* – CEEI was disappointed that the proposed EPA Guidelines did not mention a useful policy that has been under development in OEI for at least two year. Late in 2000, EPA issued a report, entitled

---

[11] As part of this process, a "flag" is placed on the data in the Website so users can know that the data are in question.

[12] Concerns about responsiveness have arisen with other Agency databases as well. In particular, companies continue to experience frustrations in making corrections to the Sector Facility Indexing Project, a database characterizing compliance status administered by EPA's Office of Enforcement and Compliance Assurance. The database continues to characterize minor violations as major violations, record inspections as violations and introduce a new round of errors when the database is refreshed. When these errors occur, transaction costs for companies can be quite high in order to make corrections. In contrast, some state efforts, such as Michigan's Community Environmental Awareness Project, have been reported to have more success in compiling accurate facility profiles.

"Lessons Learned about Designing, Developing and Disseminating Environmental Information Products," which provides a very promising policy framework for how EPA might approach data quality issues.[13]

The Lessons Learned report summarizes the collective experience of a variety of EPA, state and local officials who were involved in developing some of the major environmental information products disseminated to the public over the last several years. The document assembles those insights into a multi-step product development process that reflects "best practices." The document endorses several key aspects of what CEEI has referred to as "information products stewardship." For example, the document emphasizes the need to define the audience and purpose of a product, avoiding the imprecision that goes with viewing the "general public" as the audience. The report recognizes the need to provide contextual information to help the audience understand the product. The report also highlights the value of customer feedback, both before and after a product is released, to help refine the product and meet customer needs.[14]

The report outlines several steps that should be used to develop and maintain information products.[15] CEEI believes that EPA should explicitly adopt this process as Agency policy and begin to refine the various steps in the process. The appropriate elements of the report should be incorporated into the sections of the Guidelines that address pre-dissemination review of information. Similarly the later stages of the process should be aligned with the corrections mechanism in the Guidelines.

CEEI notes that the process outlined in the report is most appropriate for major information products, particularly those that involve significant interpretation or policymaking as part of the product's development. For simpler products, such as the posting of data submitted to the Agency without any attempt to interpret the information, the Agency may be able to complete an abbreviated version of the process. Where the Agency is meeting its obligations under the Electronic Freedom of Information Act by posting documents submitted to the Agency, the product development process outlined above may not be necessary at all.

---

[13] U.S. EPA, Lessons Learned about Designing, Developing and Disseminating Environmental Information Products (November 17, 2000).

[14] As it considers the design of its pre-dissemination review program, EPA should also consider how it will interact with the public on the development of an information product. For example, EPA should be making better use of its Information Products Bulletin as a device to alert and engage stakeholders about new information products.

[15] The steps are (1) Design – identify purpose and audience, consult with stakeholders, develop a product plan; (2) Development and Review – develop a version of the product, consult with data providers on accuracy and suitability, align product with data limitations; and (3) Dissemination, Maintenance and Feedback – release the product with appropriate metadata and context, seek out feedback from data providers and users to improve the product.

The Proposed Corrections Mechanism Is Inadequate

CEEI believes that an effective corrections mechanism is perhaps the most important component of these Guidelines. The corrections mechanism defines the critical role of the public in making sure that the objectives of the Guidelines are met.

Over the last sixty years of American administrative law, certain federal statutes have made a major difference in how agencies manage themselves and how they interact with the public. A common thread of the most influential statutes, such as the Administrative Procedure Act (APA) and the Freedom of Information Act (FOIA), is that these statutes have defined a procedural role for the public. Through these statutes, a variety of private interests, from businesses to non-governmental organizations, have pushed agencies toward greater transparency and accountability.

In contrast, administrative "reform" statutes that have relied on the executive branch to police itself or on a busy Congress to see that laws are fully implemented have been much less successful. Recent examples of such ineffective reform efforts have included the Unfunded Mandates Act and the Government Performance and Results Act.

The existence of procedural rights for interested stakeholders does not necessarily lead to governmental paralysis. Governmental agencies are highly resourceful and are usually able to internalize the values of statutes with strong procedural components. Procedural rights tend to create an incentive structure for agencies to incorporate the substantive policies of reform statutes and thereby reduce the risk of excessive procedural disputes.

The availability of procedural rights is even more important in the context of the EPA Guidelines because there are few other incentives to assure adequate implementation of data quality policies. EPA's proposed Guidelines are not very specific. In fact, the Guidelines are clearly designed to provide EPA offices with great discretion in how they will be implemented. The Guidelines also do not establish strong authorities in a central office, such as OEI.[16] The diffused leadership responsibilities in the Guidelines raise serious doubts about the accountability of EPA offices for meeting the terms of the Guidelines.

In its current form the proposed corrections mechanism is unlikely to be effective. Perhaps the most obvious failing is the absence of any time deadlines defining how long Agency staff may take to make decisions on correction requests. The OMB Guidelines are quite explicit that agencies are expected to "specify appropriate time periods for

[16] For the most part, OEI performs a ministerial role under the Guidelines. A possible exception may be OEI's role in appeals of decisions made on correction requests. As a more general concern, CEEI has expressed the view on several occasions, including the announcement of the creation of OEI, that EPA has not provided OEI with sufficient power in the Agency to assure implementation of information policy.

agency decisions on whether and how to correct the information."[17] In reviewing the guidelines that other agencies have issued, CEEI finds that many agencies have specified time deadlines for responding to correction requests.[18]

EPA should establish appropriate deadlines for decisions required under the corrections mechanism. These deadlines should match that nature of the requested action. As suggested by EPA's experience with the Integrated Error Correction Process, most of the requests for changes to data will involve fairly straightforward corrections of inaccuracies. Resolution of these requests should not be particularly time-consuming. Accordingly, EPA should be able to establish timeframes for corrections of data inaccuracies that would not exceed a few weeks. As a benchmark, EPA could consider a three week timeframe. Such a timeframe roughly corresponds to the Agency's deadlines for responding to FOIA requests or the deadlines private parties must meet for voluntary self-disclosure under EPA enforcement policies.[19]

Some correction requests will involve broader interpretive issues or technical questions. While EPA's review under the correction mechanism will remain focused on the information policy aspects of these issues, the Agency may need more time to deliberate on such requests. As a benchmark EPA could consider a 60 day time frame for such requests. EPA should be able to resolve most requests within such a timeframe. Several other agencies have incorporated such a deadline in their correction mechanisms.

EPA has also expressed some interest in defining a timeframe after which a correction request would not be allowed.[20] While CEEI understands that EPA might desire some kind of "statute of limitations" for correction requests, such a concept is not compatible with the dynamic nature of information use.[21] Older data-sets can become relevant to new "secondary uses" at any time. EPA or other parties may find new applications that raise new data quality issues that need to be addressed. In addition, new

---

[17] OMB Guidelines, at III. 3.i. This point was reinforced in the OMB Memorandum, at Section V.

[18] Some agencies have also specified time deadlines for requesters to file correction requests and appeals from initial decisions. If EPA chooses to establish such timelines, they should provide adequate time to allow submitters to assemble necessary information.

[19] U.S. EPA, Incentives for Self-Policing: Discovery, Disclosure, Correction and Prevention of Violations, 65 Fed. Reg. 19618 (April 11, 2000).

[20] EPA asked commenters at its May 15, 2002 public meeting to address this issue.

[21] The OMB Memorandum, at Section V., echoes this concern, noting that the "damaging effects of poor quality information may not occur or be perceived to have occurred until well after the information was originally disseminated."

discoveries about the quality of data underlying EPA actions can legitimately require new inquiries into data quality.[22]

A second area of concern with the correction mechanism is the criteria that EPA will use to decide when corrective action should be taken. Under Section 5.5 of the proposed Guidelines, EPA indicates that it "may elect not to correct some completed information products on a case-by-case basis due to Agency priorities, time constraints, or resources." The implication of this statement is that EPA may not proceed with meritorious correction requests. Moreover, EPA has broadly stated the conditions warranting inaction. The Agency will almost always be able to identify alternative priorities, time constraints or resource limitations that could provide a basis for ignoring a valid correction request.

CEEI believes that the decisionmaking criteria in Section 5.5 send the wrong signal on EPA's seriousness about the data quality issue. The Guidelines certainly authorize EPA employees to ignore correction requests that have any effect on day to day operations. In turn, the public cannot have confidence in a corrections process that does not reflect a serious commitment to improving data quality. EPA should make it clear in the Guidelines that corrective action concerning valid requests should take place as soon as feasible.

Section 5.5 of EPA's Guidelines should articulate the standards under which correction requests will be evaluated. The OMB Memorandum, at Section V., makes it clear that an agency must allow complainants to raise issues under all applicable guidelines, including those of the relevant agency and of OMB.

A third area of concern about the corrections mechanism is the appeal process. The Guidelines indicate that appeals will be handled initially by an "executive panel", chaired by the EPA Chief Information Officer. The executive panel makes recommendations to the responsible Assistant Administrator or Regional Administrator who makes the final decision on the appeal.

The executive panel process is unclear. Is the panel made up of senior staff from across the Agency or from the office affected by the correction request? It is also not clear what criteria would be used to select members of the panel. For example, is this an effort to bring technical expertise to bear on the questions in the appeal? Will the panel be limited to specific technical questions, as is typical in peer reviews? It is also unclear how the ultimate decisionmaker is to view the recommendations of the panel. Are the recommendations purely advisory or must the decisionmaker defer to the panel on certain questions?

---

[22] As a recent example, the Health Effects Institute has disclosed that investigators of the National Morbidity, Mortality and Air Pollution Study have determined that certain statistical default criteria used in analyses underlying EPA's air standards may have caused an overestimate of health effects from exposure to particulates. Letter from Dan Greenbaum, Health Effects Institute to colleagues (May 30, 2002).

EPA should consider whether the executive panel approach is really necessary for all appeals. The process could become unwieldy and time-consuming, particularly if EPA finds it difficult to accommodate the schedules of busy senior officials. In many cases, the issues raised by appeals will be discrete policy questions that require decisions, but not necessarily elaborate deliberations. If EPA decides to retain the executive panel process, it may make more sense to invoke the process for more complex appeals that would benefit from technical expertise.

The more fundamental difficulty with the proposed appeals process is its ability to assure objective decisionmaking. In explaining the intent of the corrections mechanism, OMB makes it clear that "an objective process will ensure that the office that originally disseminates the information does not have responsibility for both the initial response and resolution of a disagreement."[23] Assuring some distance between the office that initially responds to a correction request and the appellate office would be consistent with other similar administrative mechanisms used by agencies. For example, EPA's FOIA rules provide that initial questions about the scope of a FOIA response will be made by relevant program offices but that appeals of those decisions will be handled centrally by the Office of General Counsel.[24]

CEEI believes that appeals of correction requests can and should be made by a central authority in the Agency. Certainly appeals could be handled by the Administrator or Deputy Administrator, served by an appropriate staff dedicated to such appeals. In the alternative the appeals could be decided by the Chief Information Officer (or Assistant Administrator for Environmental Information.)

Centralization of the appeals process has clear advantages. It will provide the necessary neutrality to assure the integrity of the process. Without questioning the personal integrity of EPA's Assistant Administrators and Regional Administrators, it is institutionally difficult for people in those positions to overrule their staff on matters in dispute with external parties. Certainly parties outside the Agency have that concern.[25]

Centralization would also help provide consistency in how EPA approaches matters of information policy. EPA is a highly decentralized agency and decisions about how information is collected, used and disseminated have historically been made as secondary questions by a wide range of EPA decisionmakers. This structural and cultural framework at EPA has made it extremely difficult for the Agency to respond systematically to the challenges and responsibilities of the Information Age, in which an EPA Website can have as much practical effect as a regulation.

---

[23] 67 Fed. Reg. 8458 (February 22, 2002).

[24] Subpart A of 40 CFR Part 2.

[25] The OMB Memorandum, at Section V., emphasizes that agency appeal mechanisms should "build public confidence in both the reality and appearance of a neutral, fair decision mechanism."

In the late 1990s EPA and the Congress began to realize the importance of developing consistent information policies. This understanding led to the creation of OEI, which was to be a center of leadership in the Agency on information policy. Perhaps the greatest challenge EPA has had to face in achieving a broader vision for the strategic use of information is the consolidation of decisionmaking on information policy so that EPA can operate as one agency rather than a confederation of interests. Centralizing the appeals process in the Guidelines would be one useful step in that direction.

EPA can centralize the appeals process without interfering with other policymaking functions of the Agency. Matters of data quality necessarily overlap with science policy and regulatory policy. Yet it is feasible to clarify the aspects of a controversy that relate to data quality issues and focus the appeals process on those issues. Substantive decisions can be reserved to line program offices.

For example, if a submitter raises questions about the quality of a scientific model that is being used for regulatory purposes, the Agency can distinguish the "information policy" questions from substantive decisions about the use of the model. The Guidelines, for example, will focus on the need for transparency and clear explanations for how the model works. Implementation of the Guidelines will also consider whether the model represents the best available scientific data. The question of whether a program office should use the model to implement its statutory or regulatory authorities will ultimately be made by the program office, after consideration of the data quality issues and other factors within the scope of its legal authority. Disputes about that substantive decision will be resolved through the normal channels for agency action and judicial review.

In considering the appropriate structure for an appeals process, it is important for EPA to recognize that there are cross-cutting information policy issues that deserve consistent treatment. EPA appears concerned that the Guidelines not intrude on the prerogatives of Agency offices to make substantive policy decisions. At the same time, it is no more appropriate for individual program or Regional offices to set Agency policy on the transparency of models, the design of Websites, the description of metadata or the presentation of scientific information. EPA should operate as one agency on these matters.

A final area of concern about the corrections mechanism involves the overlap with rulemaking procedures. In Section 5.4 of the Guidelines, EPA indicates that the corrections mechanism will not be available "where a mechanism by which to submit comments to the Agency is already provided." In particular, EPA indicates that the rulemaking process will "assure a thorough response to comments on quality of information" and that a separate process for information subject to public comment "would be duplicative, burdensome, and disruptive" to the rulemaking process.

While CEEI recognizes that there will be a need to reconcile the rulemaking provisions of the APA with the information dissemination requirements of the Data Quality Act, we believe that the formulation of the issue outlined in the Guidelines is too simplistic. The rulemaking process is often very long, particularly at an agency like

EPA. The Agency cannot assume that the rulemaking process will, in all cases, be sufficient to address stakeholder concerns about the dissemination of information that occurs in the context of rulemakings. In some cases, the documents contained in rulemaking dockets have significant marketplace effects that are independent of, and considerably swifter than, any rule that EPA might issue.

A telling example of the power of information can be found in EPA's efforts to regulate asbestos under the Toxic Substances Control Act (TSCA). EPA began its rulemaking to regulate asbestos products in 1979 with the issuance of an Advance Notice of Proposed Rulemaking. In 1986, EPA issued a proposed rule that would ban all uses of asbestos, based on a finding that such use carried an unreasonable risk of injury to health or the environment. This proceeding, one of EPA's first efforts to remove a substance from commerce, was a strong signal to the marketplace. By the time the TSCA ban was proposed, many product manufacturers had taken steps to remove asbestos from their products. EPA issued a final TSCA ban on asbestos in 1989, and the rule was challenged in court. Two years later, the court determined that the rule was not properly justified and remanded the rule to the Agency.[26]

Despite the fact that the asbestos rulemaking was ultimately unsuccessful as a rule, the rulemaking had significant impact as an information dissemination. The uses of asbestos that disappeared around the time of the proposed TSCA rule, such as asbestos floor tile, did not return to the marketplace at any significant levels once the EPA rule was remanded. Product manufacturers had moved to other substitute materials for asbestos.

This example illustrates two key realities. First, EPA documents that are disseminated during the course of rulemakings are important market signals that influence private sector behavior. Those impacts can occur quite swiftly in a marketplace that depends on rapid access to on-line information. Second, the sometimes tortuous rulemaking process cannot assure a prompt consideration of data quality issues that may arise with rulemaking documents.[27]

CEEI recognizes that EPA should pursue the objectives of the Data Quality law without undercutting the integrity of APA rulemaking processes. At this stage in the development of the Data Quality law, however, it is too early to address the tensions between these laws through a sweeping, categorical exclusion of all rulemaking documents from the corrections mechanism. Instead, EPA should examine the

---

[26] Corrosion Proof Fittings v. EPA, 947 F.2d 1201 (5th Cir. 1991).

[27] In addition, at least as practiced today by agencies, the APA rulemaking process does not assure that the full set of criteria in the OMB Guidelines will be addressed in agency responses to comments about data quality concerns with particular information. Commenters on EPA rules have seen responses to their comments about flaws in information that are less than fulsome, sometimes not amounting to more than a statement that the Agency disagrees.

interaction between corrections requests and rulemaking processes on a case by case basis. Over time, it may be possible to establish some general principles that will guide how the two administrative processes should interact.[28]

The Guidelines Need to Address All Dimensions of Data Quality

OMB has indicated in its Guidelines that data quality is "an encompassing term comprising utility, objectivity, and integrity."[29] Each of these terms is defined separately. Each concept deserves equal weight in EPA's Guidelines.[30]

The EPA Guidelines dwell primarily on the "objectivity" standard, making only passing reference to the utility and integrity standards. Admittedly, the objectivity standard is perhaps the most complex of the three standard. EPA should, however, indicate in the Guidelines how it will implement the other two standards.

The "integrity" standard is concerned with the security of information, protecting agency-held data against unauthorized access, corruption or falsification.[31] During the last few years, EPA has been severely criticized by the Congress and the GAO for the inadequacy of its cyber-security measures.[32] In response EPA has undertaken a series of steps to improve its firewalls and other software systems to prevent intrusion into its computer systems. While the specifics of EPA's security measures must, of necessity, be kept confidential, it would be valuable for the Agency to identify in the Guidelines the general policies and staff responsibilities that guide the Agency's cyber-security strategy.

---

[28] The OMB Memorandum, at Section V., suggests that agencies allow for an "early response" to important data or analyses placed in rulemaking dockets when there is evidence of actual harm from dissemination of such information or when there is substantial uncertainty about the length of the rulemaking. There may be other situations where "early response" is appropriate. For example, certain critical studies, if substantially flawed, may set an agency on the wrong path that may result in unjustified agency action.

[29] OMB Guidelines, at V.1.

[30] The OMB Memorandum, at Section III., indicates that "each agency needs to adopt explicitly each aspect of each definition of quality, utility, objectivity and integrity as an agency quality standard." OMB indicates that agencies must develop their own definitions of the data quality terms, consistent with OMB's Guidelines, or explicitly adopt the OMB quality definitions.

[31] OMB Guidelines, at V.4.

[32] GAO, Information Security: Fundamental Weaknesses Place EPA Data and Operations at Risk (July 2000).

In this context, it is also important for EPA to recognize that the integrity standard can be the subject of a correction request by private parties. If a data submitter has reason to believe that the security of data held by the Agency has been, or may be, compromised, it is appropriate for the submitter to seek corrective action through the corrections mechanism.

The "utility" standard is also an extremely important aspect of the Guidelines. Under the OMB Guidelines, this standard "refers to the usefulness of the information to its intended users, including the public."[33] OMB emphasizes that agencies should consider the uses of information from the perspective of the public, rather than just internal agency audiences. The utility standard, with its focus on the needs of an intended audience, should be read in conjunction with the portion of the "objectivity" standard that calls for the presentation of data in an "accurate, clear, complete, and unbiased manner."[34]

The question of whether a particular information dissemination is clear, complete and unbiased in its presentation is best understood in terms of the intended audience for the information. If the audience does not understand the information or draws unintended conclusions from the dissemination, then neither the utility nor the objectivity standard has been met. EPA's obligation to communicate environmental information in an understandable way is an increasingly important responsibility. The Agency's Website is now a mammoth enterprise that receives tens of millions of "hits" on a monthly basis. Yet the Agency has only minimal information about how the public uses the Website.

EPA should address the utility standard more explicitly in the Guidelines and in the implementation of its data quality program. To some extent, adoption of the Information Products Development Process, discussed earlier in these comments, would provide a starting framework for effective implementation of the "utility" standard. That Process contemplates significant interaction with data submitters and users in the design and development of information products, an approach that is likely to enhance the quality of information products offered to the public. At some level, the utility standard also provides a benchmark for whether certain types of information products should be developed. A database that contains major problems of methodology or analytical rigor should not be offered to the public as a guide to understanding environmental issues.

The Guidelines should also recognize that the utility standard can provide a basis for correction requests. The utility standard is likely to be invoked when parties believe that EPA's current presentation of data leaves a misleading impression about environmental performance or environmental conditions in the public mind. Resolution of complaints of this nature is likely to be one of the most important applications of the Guidelines at EPA.

---

[33] OMB Guidelines, at V.2.

[34] OMB Guidelines, at V. 3.a.

EPA Should Improve its Approach to the Objectivity Standard

Much of the discussion surrounding the Data Quality Act has centered around the meaning of the "objectivity" standard. Under the OMB Guidelines, objectivity "involves a focus on ensuring accurate, reliable and unbiased information."[35] It should be noted that this general standard applies to all EPA information, not just information that is deemed "influential" as discussed below.[36]

EPA's Guidelines devote an entire section to defining the "influential" information that is subject to the reproducibility standard.[37] CEEI believes that the list of influential information should be broader. This modification is necessary to fully implement the OMB Guidelines.

Under the OMB Guidelines, "influential" data includes information that has a "clear and substantial impact on important public policies or important private sector decisions." This definition capture two important, but distinct concepts. Some information is important because it shapes public policy decisions by governmental agencies. Other information, which is not necessarily critical to governmental decisions, may have a strong influence on private sector decisions and thereby be quite influential. This latter role needs to be more clearly recognized in the Guidelines. It is, in many ways, the essence of the "right to know" principle that EPA has been promoting since the late 1980s.

The EPA Guidelines do not clearly recognize this "private sector impact" component in the definition of "influential" information. The information identified in Section 3.2 of the Guidelines relates to traditional governmental policy actions such as rules. Even the references to major scientific and technical work products is limited to products "used in decisionmaking", implicitly governmental decisionmaking.

The Guidelines should explicitly recognize that information which affects important private sector decisions is "influential" information. The fact that information which EPA views as influential for other reasons may also influence private sector choices is not a sufficient recognition of the issue.

In providing greater clarity on this point, EPA may assume that an information product explicitly discussing the environmental performance of particular facilities, companies or products is likely to be viewed as influential. Certainly efforts to rate

---

[35] OMB Guidelines, at V. 3.b. As indicated earlier in these comments, the objectivity standard also encompasses assuring accurate, clear, complete and unbiased presentation of information.

[36] "Influential" information is required to be sufficiently transparent to "facilitate the reproducibility of such information by qualified third parties." OMB Guidelines, at V. 3.b.ii.

[37] EPA Guidelines, at 3.2.

companies on environmental protection criteria or to influence consumer choices through publication of "environmentally preferable product" lists would be viewed as influential information. Similarly information products that assign environmental "scores" to particular facilities will also be viewed as influential.

In some cases, the description of environmental conditions in communities could become influential. As an example, EPA is continuing to refine its GIS capabilities to describe environmental conditions in specific locations, taking the environmental information down to the neighborhood level. At some point, such information could begin to influence local real estate decisions, potentially causing "red-lining" of neighborhoods that are viewed as environmentally undesirable. If such a phenomenon were to develop, the relevant EPA data could become influential.

The latter example demonstrates an important aspect of the meaning of "influential" under the Guidelines. As data precision increases and new applications of EPA data arise, information may become influential. It is difficult to anticipate all of the ways that environmental data may influence private sector decisions. Thus there is a critical link between the corrections mechanism and this definition of influential information. Private parties are not likely to request corrections of data that have no influence. As part of a correction request, companies will often articulate why data is influential in their particular situation. Where companies present cogent explanations of why particular data is influential, EPA should be reviewing the data under the appropriate standard for influential data.

In describing the general reproducibility standard, the EPA Guidelines incorporate, for the most part, the general language found in the OMB Guidelines. EPA should consider amplifying the discussion as it relates to the use of proprietary databases, a growing trend within EPA programs. The reproducibility standard draws an important distinction between original data and analytical results. For original data there may be more situations where ethical, confidentiality or feasibility factors make it difficult to provide full transparency of the information, and thereby facilitate reproducibility.

In the case of analytical tools, such as proprietary models used in decisionmaking, it is important that agencies push for maximum transparency of the methodologies, assumptions, defaults and algorithms that underlie the model. This is particularly true when technical models are used as the primary basis for setting regulatory standards. The business community has become increasingly concerned in the last several years that some of EPA's most important regulatory decisions are based on these "black box" proprietary models.

The OMB Guidelines indicate that agencies should conduct "robustness checks" of analytical results when the parameters of the analysis cannot be disclosed. Before adopting that approach, however, EPA should be asking hard questions of the owners of proprietary models, particularly when those owners are EPA contractors who are supporting EPA rulemakings, about what aspects of their models can be made public. In addition, EPA should be pursuing a general policy that favors the use of fully transparent technical models over proprietary models of similar analytical quality as a means to

foster public understanding and to comply with the Data Quality Act. CEEI urges the Agency to send a strong signal to the marketplace of contractors seeking government work that transparency is the preferred approach.

Finally, EPA should embrace more explicitly the quality principles articulated in the Safe Drinking Water Act (SDWA). Under the OMB Guidelines, agencies are expected to "adopt or adapt" these principles in relation to "analysis of risks to human health, safety and the environment."[38] In its proposed Guidelines, EPA indicates that it intends to "adapt", rather than "adopt", the SDWA principles. The Guidelines also appear to limit the SDWA principles to the presentation of information about human health effects, rather than ecological risk or safety assessments.

While the distinction between "adopting" or "adapting" a set of principles can be a fairly subtle theoretical question, the real issue that EPA should confront is why it would not apply the principles of one of its own statutes as broadly as possible in the Agency. If any agency is in a position to "adopt" the SDWA principles, it would be EPA.

The principles found in 42 USC §300g-1(b)(3)(A) & (B) of the SDWA are reasonable, common sense concepts of good risk assessment and risk communication:

- Agencies should use best available, peer-reviewed science conducted in accord with sound and objective scientific practice;
- Agencies should use data collected by accepted or best available methods;
- In public risk documents, agencies should provide information about populations affected by a risk;
- Public documents concerning risk should articulate the range of risks and the expected risk (i.e., central tendency of available data);
- Public documents concerning risk should articulate the uncertainties associated with the risk and identify research that might reduce uncertainty; and
- Public documents concerning risk should identify peer-reviewed studies that draw differing conclusions about the risk.

These principles all reflect good practices that EPA professionals and the broader scientific community have supported for years.

Whether it is called "adoption" or "adaptation" of these principles, we cannot understand why EPA would resist these principles as a general matter. CEEI also does not understand why EPA would try to limit their applicability to human health assessments. These principles have general applicability to all manner of scientific risk assessments, including those addressing ecological or safety concerns.[39] There may be

---

[38] OMB Guidelines, at V. 3.b.ii.C.

[39] It is not clear in the Guidelines or the preamble what EPA means in referring to "safety" assessments. Depending on what EPA has in mind, such assessments might be synonymous with health assessments.

unique situations, perhaps due to emergency conditions, in which this level of documentation is not feasible. Targeted exceptions may be warranted. At this point, however, it is important for EPA to establish the general rule – that the Agency will follow the reasonable scientific principles contained in one of its own statutes.[40]

Accordingly, CEEI recommends that EPA embrace the SDWA principles in all of its risk assessment activities, including the presentation of scientific information to the public. The burden of proof should be on those EPA offices who believe that these principles are inapplicable or infeasible for their operations.

Applicability of the Guidelines to Third Parties

EPA has addressed how the Guidelines will apply to third party data.[41] CEEI believes that the discussion in Section 1.2 explaining when the Agency has "disseminated" information establishes the right principles. If EPA distributes information "in a manner that reasonably suggests that EPA endorses or agrees with it, if EPA indicates in its distribution that the information supports or represents EPA's viewpoint, or if EPA in its distribution proposes to use or uses the information to formulate or support a regulation, guidance, policy or other Agency decision or position," then the information would be subject to the Guidelines.

As indicated in Section 1.2, EPA can use disclaimers to clarify the status of information to the public in order to manage the question of whether particular information is an Agency dissemination. EPA routinely receives and posts on the Internet various documents that would be viewed as the opinion of the submitter but not necessarily the position of the Agency. Such posting may well be required to comply with the Electronic FOIA statute. Examples of such submissions include petitions for administrative action or submissions in the context of voluntary programs.

This approach requires submitters to decide at the time of submission how they would like the information treated at the Agency. If a third party hopes to convince EPA to adopt its viewpoint, the data submitted should be developed and presented consistent with the Guidelines. If the company does not meet the Guidelines, then the third party cannot expect that EPA can rely on the data without further work.

---

[40] EPA may want to consider how the Food and Drug Administration (FDA) has adapted the SDWA standards. FDA draws a distinction between quantified and unquantified risk assessments. The former category of assessments is subject to the full set of SDWA principles, while the latter category is subject to the core principles that logically would apply to all situations.

[41] References to this issue can be found in Sections 1.2, 1.3 and 3.5. It is not always clear whether EPA is attempting to resolve this question or simply reserving the issue for another day in some of these discussions.

CEEI believes that these same principles are applicable to state-generated data, an issue that EPA has not yet fully addressed. When EPA disseminates state-collected data as its own, the Guidelines should apply. For example, national program databases and EPA Web resources such as Envirofacts present the "official file" of EPA on particular facilities. The data in those files come primarily from state data collection activities. There is little question that the Guidelines should apply to such data.

This result is not simply an artifact of the Data Quality Act. It is also the logical extension of the structure of federal environmental law. States have a central role in the implementation of federal programs. Under most EPA statutes, states are authorized to operate "in lieu of" federal authorities to implement federal law. The states accept large amounts of federal money to implement those programs. The data collected by the states are essential to implementation of federal law. The states must accept the public accountability that goes with their central role in EPA programs.

If EPA finds that the states are unwilling to take responsibility for implementation of the Guidelines, the Agency will need to take strong action. For example, EPA would need to consider making compliance with the Guidelines an explicit condition for receiving federal grants. At a minimum, EPA would have to be more explicit on its Website and in other accessible information sources that state submissions which do not accord with the Guidelines do not represent the views of EPA and cannot be deemed the official file for purposes of federal programs.

EPA Should Avoid Overly Broad Exemptions

Section 1.3 of the Guidelines enumerates a long list of information sources that are exempt from the terms of the Guidelines. This section is the longest section in the Guidelines. While some of the exemptions in this section are reasonable, CEEI believes that many of the exemptions have been drawn too broadly. The most important cross-cutting flaw in this section is that EPA has defined many of the exemptions in terms of the <u>form</u> of the document or communication, rather than its <u>intent</u> or <u>content.</u>

In some contexts, however, any of these different forms of communication could actually be Agency disseminations that articulate positions to broader audiences. EPA has a long history of communicating major positions through a variety of vehicles, including guidance documents, question and answer documents and letters. It is quite common for EPA to provide major statements of position, including the characterization of important scientific information, in letters that are sent to private parties, to other agencies or to EPA Regional offices. Sometimes EPA intends to distribute these position statements to a broader audience at the time they are created. In other cases, such letters evolve into broader statements of position over time as EPA incrementally endorses the position.

Similarly a press release can be accompanied by a "fact sheet" or other supplementary material that can constitute a major dissemination of information to the

public.[42] Reports to Congress would also be considered major information dissemination activities, even though such reports could take the form of a letter to an individual member of Congress.[43]

CEEI recognizes that certain types of routine communications with parties inside or outside EPA would not necessarily be viewed as a significant information dissemination. The fact that a communication is sent as a letter to a single individual could certainly be one factor in deciding whether a significant information dissemination has occurred. Yet the form of the communication cannot be the single deciding factor. EPA should look to the intent of the communication and the effect of the communication before concluding that the Guidelines should not apply.

EPA should also take a closer look at its proposed exclusion for outdated information. The Guidelines exempt "outdated or superseded EPA information that is provided as background information but no longer reflects EPA policy or influences EPA decisions." In this situation, EPA proposes to indicate, through a disclaimer statement, that the superseded information is made available for background purposes only and does not reflect EPA's position.

The treatment of superseded information is an issue of increasing significance as Internet Websites begin to age. Agencies like EPA often maintain old documents on their Website, even though those documents do not reflect current positions. Web browsers are not designed to distinguish older and newer Web pages. Many Web pages, particularly older ones, do not indicate when they were generated. The commonly used browsers are not able to distinguish between current Agency positions and superseded information.

As a result, a search for a common environmental term, such as the name of a chemical of concern, can turn up a wide array of documents that are no longer valid. Moreover, the protocols for assembling Web pages in a priority order do not always reflect the currency of the document. The first set of documents listed in a browser may include superseded information.

EPA needs to institute clearer policies in this area. As a general matter, EPA should be removing superseded information from its Website to reduce the potential for confusion. CEEI recognizes that there will be situations where out of date information

---

[42] It is possible to read the Guidelines as limiting the "press release exemption" to situations where EPA has already disseminated the underlying information through some other means. Certainly the OMB Memorandum, at Section I., interpreted EPA's intent in that manner. We ask that EPA clarify that its position is in accord with the OMB Memorandum.

[43] As with press releases, EPA should also caveat the "submissions to Congress exemption" as only applying when the Agency has disseminated the underlying information in the Congressional communication through some other means.

may be valuable as "background" information. The value of such information, however, should not be presumed. Instead, EPA should ask hard questions about whether superseded information really serves a clear "background" purpose for a specific audience. Unless superseded information serves a very clear objective, it should be removed from the Website in the interest of avoiding misinterpretations. If superseded information is retained on the Website, EPA should establish a clearly stated disclaimer that appears, preferably as a "pop-up screen", on each page of the document.

Finally, EPA has drawn the exemption for "subpoenas or adjudicative processes" too broadly. CEEI understands that EPA would want to avoid applying the Guidelines to every piece of paper ever filed in a formal adjudicatory process or in a court proceeding. At the same time, there will be times when documents that are related to such proceedings have independent effects that warrant application of the Guidelines. As indicated earlier in these comments, EPA should be looking to the intent and effect of particular documents before deciding whether the Guidelines should apply.

A few examples indicate why a generic exemption of this nature is inappropriate. In the context of a major litigation, responsible officials may prepare and file affidavits providing factual characterizations or articulating Agency positions that may never have been announced before. These statement may constitute a major information dissemination for purposes of the immediate issue before the adjudicatory tribunal or for broader purposes outside the proceeding.

Another primary example arises in the pesticide program. The Federal Insecticide, Fungicide and Rodenticide Act (FIFRA), the primary statute under which EPA regulates pesticides, is primarily a licensing statute. Thus the documents generated in the administration of this program are arguably documents related to the issuance of a license, or cancellation of a license, to sell and distribute pesticides in this country. In the last several years, particularly as the Agency has used FIFRA to implement the provisions of the Food Quality Protection Act, EPA has been generating a series of major documents that assess the risks of particular pesticides and propose EPA's preferred risk management strategy.[44]

These major assessment documents clearly meet EPA's criteria for being influential information that should be subject to the full range of considerations set forth in the Guidelines. Yet these documents, which are often used by EPA to push pesticide registrants toward "voluntary" risk management measures, can also be characterized as preliminary documents in the FIFRA licensing process. EPA's broad exemption for "adjudicatory processes" makes reference to licensing proceedings and thus might be interpreted to exclude all of these major pesticide documents from the Guidelines. Under any theory of what the Guidelines should cover, such a result would not make sense.

---

[44] These documents take several forms, including Reregistration Eligibility Decisions, Interim Reregistration Eligibility Decisions, and Tolerance Reassessment Progress and Interim Risk Management Decisions.

As with the overlap with rulemaking, discussed earlier in these comments, EPA should not try to establish a broad exemption for adjudications at this time. EPA should consider the interplay between the Data Quality Act and other laws affecting administrative proceedings on a case by case basis. Over time it may be possible to craft more specific policies on this issue.

EPA Should Establish a Management System for Pre-dissemination Review

An important aspect of the Guidelines is the program for pre-dissemination review of information resources. Unless there is an effective management system that occurs before new information products are placed in the public domain, many of the objectives of the Data Quality Act will not be achieved. Yet the discussion of pre-dissemination review, in Section 4.1 of the Guidelines, is very brief and simply recognizes that EPA offices will develop their own procedures in this regard.[45]

CEEI believes that something more is needed. As discussed earlier in these comments, we believe that the Information Product Development Process would provide a sound starting point for a useful Agency-wide policy.

We would note, however, that it is important for EPA to classify the types of information it manages. Not all aspects of EPA's information management activities would need to be subject to the Information Product Development Process. As discussed in more detail in the attached "Data Quality Challenge" paper, we believe that it is helpful to divide EPA's information management activities into three categories:

- Management of original documents – EPA receives and generates many documents that must meet legal requirements about public access and security.

- Compilation of data from other documentary materials – EPA assembles data from documents into interactive databases used by a variety of internal and external audiences.

- Development of interpretive information products – EPA develops information products for the public that interpret, characterize or explain environmentally significant information.

---

[45] The OMB Memorandum, at Section III., emphasizes that a recitation of existing policies, without clear performance goals, does not meet the requirements of the OMB Guidelines. EPA should also explain how it will incorporate data quality in the core processes of the Agency. For example, the OMB Memorandum, at Section IV., indicates that agencies should demonstrate in their draft information collection requests under the Paperwork Reduction Act how the information to be collected will satisfy the data quality criteria.

The Information Product Development Process, discussed earlier, is most appropriate for this third category of activities. The Process has little applicability to the management of original documents, where EPA's primary responsibilities are defined by a variety of other statutes, including FOIA. The second category of activities does require some consideration of how information should be presented to address data quality concerns. For example, understandable metadata that explain the strengths and weaknesses of data-sets are needed. In addition, there must be an effective error corrections process for major databases. EPA should consider how the Information Product Development Process can be adapted for the development of such databases.

Effective Date of the Guidelines

The OMB Guidelines indicate that the pre-dissemination review elements of an agency's guidelines should apply to information that the agency first disseminates after October 1, 2002.[46] The corrections mechanism applies to information that an agency disseminates after October 1, 2002, through a Website or other vehicle, regardless of when that information was first disseminated. EPA's Guidelines should incorporate OMB's general policies in this regard.

In considering how the effective date of these Guidelines relates to the corrections mechanism, the OMB Memorandum, at Section I., suggests that agencies can consider exemptions for "archival records." At the same time, the OMB Memorandum clearly states that the Guidelines apply to "existing official agency data bases, publicly available through agency websites or other means, that serve agency program responsibilities and/or are relied upon by the public as official government data." In the context of EPA, major Websites such as Envirofacts, IRIS, Index of Watershed Indicators or the National Air Toxics Assessment fall into the latter category. The fact that a particular file is old does not make it an archival record, particularly if the file remains an official policy position of EPA.

Conclusions

As indicated in these comments, CEEI believes that the EPA Guidelines need improvement to bring them in line with the Data Quality Act and with the OMB Guidelines. We also believe that EPA will need to address some EPA-specific questions that are important to assure effective implementation of a data quality program.

We hope that EPA will view implementation of the Data Quality Act as an opportunity. EPA has been at the forefront of information policy issues in the federal government for many years. The Agency's ambitious efforts to use its information resources strategically are also well-known. If EPA intends to maintain its leadership

[46] OMB Guidelines, at III.4.

position and to find new ways to serve public needs with information, it must actively pursue a strong data quality agenda. A perception that EPA is not serious about data quality issues undermines EPA's desire to be a credible source of information for the public. CEEI urges EPA to step up to the data quality challenge.

Coalition for Effective Environmental Information
June 14, 2002

7124982

Attachment 1

THE DATA QUALITY CHALLENGE

I. Executive Summary

The federal government is using the tools of the Information Age to advance its objectives for environmental protection. These activities have raised a variety of new issues about how information should be collected, used and disseminated. One of the most important of these issues is the need for data quality of the information provided to the public.

Few stakeholders would argue with the basic premise that the government should be providing the public with an accurate picture about what is important on environmental issues. Yet debates have arisen about data quality, driven by disagreements over the meaning of data quality and over what steps would be necessary to deliver high quality data.

This paper is intended to assist public dialogue about data quality issues by:

1. Identifying the key dimensions of data quality;
2. Analyzing the root causes of disputes about data quality; and
3. Offering a framework for addressing data quality issues.

Data quality is best thought of as a set of attributes or characteristics that cause information to have value to its users. These "dimensions" of data quality take many forms:

- *Accuracy* - As a rudimentary matter, the government should convey to the public the information it receives without introducing errors. Some groups would include concerns about the quality of information submitted to the government within the "accuracy" label as well.
- *Scientific integrity* - Some people view out of date, or less rigorous, scientific assessments as raising a data quality concern.
- *Statistical validity* - Some groups have criticized the government's inclination to collect and disseminate data without designing a statistically valid approach for collecting information that can answer a question with a known level of confidence.
- *Comparability of measurements* - Depending on the context, data quality can also include the ability to make reliable comparisons of data across geographic lines or over a set time period.
- *Data integration* - Some people see the integration of data to capture the full profile of a place or activity as an essential element of data quality.

- *Perceived data gaps* - For some groups the absence, or lack of availability, of valuable data is seen as a breakdown in data quality.
- *Timeliness* - Depending on an information user's needs, the age of data can become very important.
- *Contextual information* - Data users often seek additional information to understand data presented to them. This need may include information about how the particular data at issue was collected, or may require supplemental information about the significance of particular data.
- *Relevance to an audience* - Questions about data quality also arise when a database is not designed to answer the particular questions of interest to data users.

All of the above concerns are legitimate dimensions of data quality. Certainly they define whether or not a particular information user finds value in the data. Policies on information management should be designed to advance all of these values.

While differing conceptions of the meaning of data quality can foster controversy, there are other "root causes" of disputes in this area that have become apparent in recent years. These root causes of disagreement include the following:

- *Transparency vs. Accuracy* - One of the primary reasons for public disclosure laws is to create incentives for government agencies or private parties to conduct themselves responsibly. From this perspective, the accuracy of information disclosed is important, but not necessarily the primary objective. Those who see the act of disclosure as the primary objective have come into conflict with those who place a higher value on assuring that information is correct before it is disclosed.

- *Attempting to Serve Multiple Purposes* - EPA often develops information resources that try to serve many purposes. Depending on the user's needs, however, the various dimensions of data quality become more or less important. For example, some level of inaccuracy in individual data points can be acceptable when analyzing broad national trends, yet such inaccuracies are unacceptable when describing environmental performance at particular facilities in a community.

- *Attempting to Serve Multiple Audiences* - At times EPA has developed information resources for new audiences with data that was collected with another audience in mind. For example, EPA and state regulators, who share data on facility compliance with regulations or on broad ambient conditions, recognize the limits of such data. More recently, however, these data are also being used to create interactive information resources for the general public about the state of local environmental quality. Both government and private sector groups have become concerned that the public will not appreciate the limitations of the underlying data.

- *Maintaining Objectivity in Public Discourse* - Government agencies are expected to present their own information in a balanced and objective manner. Third parties routinely use and

interpret government data to communicate their viewpoints on environmental issues. At times these third party characterizations have become controversial, and EPA has been asked to issue public statements to clarify the record. The conditions under which EPA should take on this role is a source of significant debate.

- *Improving vs. Describing Data Quality* - EPA has issued several public statements indicating that it intends to enhance the quality of the information it disseminates to the public. For some purposes, however, it is probably necessary to collect different information than the Agency currently collects in order to improve data quality. For example, facility operations data do not offer strong insights into the state of public health or ecosystem vitality. Yet, EPA's efforts to "improve" data quality have historically been limited to better descriptions of the data that the Agency currently collects.

The factors described above have made data quality somewhat controversial, despite the fact that data quality, as a general concept, is highly valued by most participants in environmental information issues. Greater consensus could be achieved if these parties shared a common framework for how EPA, and other agencies, could define and address data quality needs. This paper recommends the following management framework:

- EPA is engaged in three general categories of information management activity that carry differing data quality responsibilities. First, EPA is responsible for managing original documents that contain environmental data. These documents may be generated within the Agency or submitted by outside parties. Second, EPA compiles data from original source documents (e.g., forms, letters, reports, permits) into interactive databases that can queried by a variety of users. Such data compilations are generally designed to serve the analytical needs of an audience within or outside the government. Third, EPA develops interpretive information products that characterize environmental performance or conditions.

- In its management of original documents, EPA's responsibilities equate to its FOIA obligations. The Agency should provide timely and complete access to documents of broad public interest. EPA should clarify the source and origin of documents when that information is not clear on the face of the document.

- In compiling databases to facilitate research, EPA should meet several data quality objectives. First, the database should track accurately the data contained in the original source documents. Second, the database should have a clear and understandable explanation of the parameters for the data's collection (e.g., age, methodology, scope). Third, EPA should provide a general definition of the appropriate use of the data.

- When formulating an interpretive information product, EPA should implement the Information Stewardship Principles, explained in section V.C. The starting point for application of the Principles is the need to define the purpose and audience for the information product. Once the purpose and audience are defined, EPA can establish its expectations for the dimensions of data quality regarding accuracy, timeliness, comparability, precision, contextual information, data needs and procedural steps. These dimensions of data

quality, grounded in the product's original purpose and audience, become the key criteria used in the design, implementation and periodic evaluation of the product.

- There will be times when it is difficult to place an information resource in the three-part classification scheme outlined above. In particular, this can occur when distinguishing a compilation of data from an information product. EPA should apply a functional test (i.e., how does the public understand and use the data) when deciding whether to apply the Stewardship Principles to a particular information resource.

II. Introduction

With the arrival of the Information Age, the public has been provided on-line access to a tremendous volume of environmental information. This information is provided by government Web sites, such as the on-line offerings of EPA, as well as a variety of private sector sites.[1] Many of these information sources are receiving heavy use by the public. The future promises growing use of these resources as the concept of on-line access to environmental information moves from a novelty to an expectation.

Institutions that aspire to be primary information providers for the public necessarily take on a set of new policy and management issues. The nature of these new roles and responsibilities has been the topic of significant discussions over the last several years. While there are disagreements on some areas of responsibility, there is a broad consensus that all information providers should be committed to data quality.

Providing an accurate picture about what is important on environmental issues is perhaps the core duty for any party who seeks to educate or inform the public. The data quality commitment is an essential public trust for government because the public has an expectation of objectivity and neutrality in the information it receives from public agencies, at all levels of government. Increasingly, governmental institutions have recognized that they must assure data quality in the information resources they offer the public. For example, EPA has indicated that data quality improvement is one of its primary goals.[2]

Yet there are continuing concerns about the quality of the environmental data that EPA and other parties are offering to the public. These concerns have been identified by EPA, the

---

[1]Such sites have been offered by environmental groups (e.g., www.scorecard.org), commercial groups (e.g., www.ecomall.com) and academic institutions (e.g., www.ciesin.org).

[2]The mission statement of EPA's Office of Environmental Information (OEI) includes a commitment to "provide leadership for improving quality and utility of EPA's data" and "ensure that EPA shares environmental data and information with our partners and the public in a consistent, efficient manner that avoids conflicting or confusing messages and promotes user understanding through contextual information and tools for evaluation."

General Accounting Office and various stakeholder groups.[3] Thus, despite the recognized importance of data quality and a strong expression of intent to improve data quality, government agencies face a perception that progress is not being made.

One of the primary reasons for this perceived gap between the government's commitment and performance on data quality is that various stakeholder groups have different expectations in mind when they use an encompassing term like "data quality." As a result, the recommendations for reform from various stakeholder groups take different forms. EPA has not been able to address most of these recommendations and thus has been perceived as ineffective.

In addition, some stakeholders have seen data quality measures as potential obstacles to public access objectives. These groups fear that a focus on data quality at EPA could lead to unacceptable delays in the public release of data.

The Coalition for Effective Environmental Information (CEEI) believes that a commitment to data quality is in the interest of all parties who generate or use environmental data.[4] CEEI also recognizes that there should be a common understanding of the challenges inherent in data quality as a starting point for discussions. To that end, this paper is intended to serve several purposes:

1. Identify the key dimensions of data quality;
2. Analyze the root causes of disputes about data quality; and
3. Offer a framework for addressing data quality issues.

III. The Dimensions of Data Quality

Data quality is best thought of as a set of attributes or characteristics that cause information to have value to data users. In this paper CEEI refers to these attributes as the "dimensions" of data quality. The dimensions of data quality take many forms:

1. *Accuracy* - Perhaps the most broadly accepted concept of quality is that data must be accurate. Yet even with such a basic concept, stakeholders emphasize different aspects of accuracy. At a minimum, accuracy means that the government disseminates the data it receives

---

[3]As an example, see U.S. GAO, *Environmental Information: EPA is Taking Steps to Improve Information Management but Challenges Remain*, GAO/RCED-99-261 (September 1999), p. 10.

[4]CEEI is a group of companies and business organizations committed to improving information policies that affect the collection, management, use and dissemination of environmental information. The group includes representatives from the aerospace, chemical, energy, automobile, pharmaceutical, forest products, electronics and consumer products industries.

from business or other sources without error. This concern has been a topic of particular interest in the regulated community because of the data entry errors and data transfer errors that have been found in databases that EPA has posted on-line.

Other groups have emphasized that accuracy also includes the responsibility of data submitters to convey to the government data that accurately portray environmental conditions or performance. The "accuracy" label has also been used in conjunction with concerns about the scientific methodology used to describe the health or environmental effect of a human activity or a chemical.

2. *Scientific integrity* - Data quality issues also can overlap with concerns about the objectivity of a scientific characterization. Some parties see hazard or risk assessments that do not reflect new data or new insights in toxicology or ecology as constituting poor quality information.

3. *Statistical validity* - The issue of data quality can also include questions about the statistical validity of a conclusion drawn from data. In the normal course of its information collection and dissemination activities, EPA does not design its systems to answer specific questions with a known level of statistical precision. Some stakeholder groups have criticized EPA for this approach.

4. *Comparability of measurements* - A common scenario for data quality debates arises around the need for consistency in measuring events over time and space. Data sets collected in different parts of the country may be conducted by parties (e.g., state agencies) who use differing monitoring methods or who judge the data by different standards. In addition, differences in data collection methodology can occur over time, prompted by factors ranging from method improvements to changes in funding.

Differences in measurement can also be ingrained in the design of environmental programs. For example, a project like the Toxic Release Inventory (TRI) under the Emergency Planning and Community Right to Know Act relies on annual estimates of releases from data submitters, without specifying any particular monitoring techniques, while an NPDES permit under the Clean Water Act defines specific protocols to be used in measuring discharges. Program differences can become especially important when the relevant data describes behavior or makes a legal determination. For example, the meaning of a term like "violation" in a database on compliance can have very different meanings between states or even between programs in the same agency.

5. *Data integration* - For some stakeholders a term like "data quality" implies that environmental data will be integrated to capture the full profile (e.g., at a facility, product or industry level) of an entity's impact on the environment. From this perspective, multiple separate databases that capture portions of that performance, perhaps in overlapping ways, are seen as poor quality data.

6. *Perceived data gaps* - Some stakeholders see the absence of data as a data quality problem because such a situation prevents certain inquiries or understanding. At various times stakeholders have expressed concerns about the quality of available information because of the absence of data on chemical hazards, exposure to chemicals, materials accounting data, product benefits, energy use, resource depletion, economic impacts or alternative technologies.

7. *Timeliness* - Depending on a user's needs, it can be important to know when data was collected because of the possibility that events have changed. While more recent, and even real-time, data tends to be preferred in most situations, there are times when older data is adequate for the user's purpose.

8. *Contextual information* - While the "need for context" is often cited as an element of data quality, various parties have often seen that need differently. For some, context is provided by supplying "metadata", a set of descriptions about how the data was collected and what its limitations might be. Others see the duty to provide "context" more broadly, including an educational function of explaining the significance of particular environmental data. Still others see the "context" duty as requiring an airing of the range of expert opinion about a particular environmental question so that the public can appreciate where there is and is not broad scientific consensus.

9. *Relevance to audience* - A final aspect of data quality is its relationship to an audience. From this perspective, an information resource is not really of high quality if it cannot meet the needs of its users. For example, users of a database may see it as a low quality resource if it is not designed to answer the user's questions or if it presents data that is too complex. An information resource can also be too simplistic or generalized to answer the questions of more expert audiences.

Conclusion: All of the above concerns should be viewed as valid dimensions of data quality. Certainly the concerns and values described above are relevant to whether particular information will be useful in resolving questions of interest. While some of these values might be better placed under the rubric of another term of art, CEEI does not see great policy value in excluding particular concerns from the "data quality" category. It is more valuable to define a framework that will allow all of these values to be served. Accordingly, data quality should be seen as an encompassing term that captures at least the dimensions described above.

IV. Root Causes of Disputes Over Data Quality

As indicated above, the broad range of definitions for data quality can lead to disagreements among groups about the right strategy for improving public information resources. This goal is made even more challenging, however, by a deeper set of information management issues that lie at the heart of several recent controversies about data quality. This section describes these additional factors.

1. *Transparency vs. Accuracy* - To some extent the objectives behind public disclosure laws and policies are not primarily concerned with providing correct answers to questions. The

information disclosure laws in this country, perhaps best exemplified by the Freedom of Information Act (FOIA), are designed to make sure that the government discloses the documents it holds. The impetus for these obligations is to assure that the government conducts itself in a responsible manner. Government officials who must "operate in a fishbowl" in a democratic society are more likely to operate in an ethical manner that serves the public interest.

When the primary objective of a disclosure law is to ensure open government deliberations, it is less important whether the information in the government's file is correct. There is, after all, a value in finding out that the government is actually holding inaccurate information. In the FOIA context, for example, the government is obligated to disclose the documents that it has. FOIA does not require that the government maintain any particular level of accuracy or provide any contextual information for the documents in its files.

A similar theme runs through the policy rationales offered for a "right to know" ethic in environmental policy. Advocates of the "power of the spotlight" see disclosure as a means to encourage better environmental stewardship. The operating assumption is that the fear of disclosing embarrassing information will cause private companies to upgrade their operations. Supporters of the TRI program, for example, have often highlighted its role as an incentive for better environmental management.

From this perspective, the accuracy of information that is disclosed is an important value, but not necessarily the primary objective. The act of disclosure is seen as the central goal. Public disclosure mandates for private companies are seen as serving many goals, including an incentive to improve the accuracy of data in the companies' files. For this reason, advocates of "right to know" will often oppose quality assurance steps in companies or in government programs because they see such efforts as delaying mechanisms that frustrate some of the broader goals of public disclosure.

This perspective was evident in the early stages of EPA's efforts to provide public access to environmental information databases through the Web site called Envirofacts. Several senior managers of Envirofacts were quoted publicly as saying that it was desirable to post the data held in EPA program databases, even though some of the data was likely to be wrong. [5] These officials indicated that a benefit of public disclosure was that the adverse reaction to inaccurate public disclosures would lead to improvements in the quality of the underlying databases.

The other perspective on this issue, however, is that the accuracy of the information is the principal concern. While the timeliness of data is certainly considered important, many users and submitters of information would rather have the government take the time necessary to assure the accuracy of information placed in the public domain. The Privacy Act embodies this preference, as it enables individuals to correct data about them held by the government. For

---

[5]Peter Fairley, "EPA's Internet Leaves Industry Feeling Exposed," Chemical Week (April 2, 1997), at 43.

many information users, moreover, accuracy is a rudimentary aspect of the value of any data. For data submitters, it is important for the government to "do it right the first time" as a matter of basic fairness. It can often be impossible, or at least expensive, to correct misperceptions in the marketplace that flow from the release of inaccurate data.

2. *Attempting to Serve Multiple Purposes* - One of the patterns in EPA's information offerings of the last several years is the tendency to develop Web resources that try to serve many purposes. These objectives can run from "wholesale" monitoring of national or regional environmental trends, as a planning guide for government programs, to more "retail" educational efforts that intend to inform citizens about specific conditions in their neighborhoods.

The problem with this strategy is that different objectives require different levels of data quality. The "wholesale" analysis of broad trends can tolerate some inaccuracy in individual data points, on the theory that underestimates and overestimates will tend to wash out in the broader analysis. In contrast, a "retail" project that professes to tell citizens what is occurring at a specific facility in their locality needs to assure a high level of precision and timeliness. It is very difficult for any database to serve well a wide range of purposes.

This is particularly true when EPA tries to use its databases for purposes outside of their original intent. Most of the data EPA collects was assembled to help EPA and the states monitor compliance with regulatory standards, usually technology-based requirements. EPA and the states also collect ambient air and water data, at dispersed locations, that provides a snapshot of average regional conditions. When EPA tries to use these data to draw broad conclusions about the state of the environment or to "empower" citizens to make highly local decisions, the data is seldom well-suited to these secondary purposes. Such uses naturally engender public debates about the quality of EPA's data.

A current example of this problem has arisen with EPA's efforts to evaluate the Sector Facility Indexing Project (SFIP), an initiative started almost five years ago that resulted in a public database characterizing the compliance performance of facilities in five industrial sectors. The SFIP was controversial for several reasons, and data quality was a major concern raised by the industry participants and the states.

In December of 1999 EPA completed a "project evaluation" report on the SFIP that addressed many of the concerns that had arisen about the program, including the question of data quality. The Agency's discussion of the data quality issue in the report reflects its misunderstanding of the concerns that the business community had raised.

In the evaluation EPA summarized some of the data accuracy problems that had arisen in the project.[6] For example, companies had raised questions about the description of 64 of 376

---

[6]U.S.EPA, "Sector Facility Indexing Project Evaluation", EPA Contract No. 68-W6-0021 (December 10, 1999), at 36 ("Evaluation").

enforcement actions in the database, and EPA accepted the comments on 41 of the actions. Similarly, EPA received requests for corrections on 3,000 of the possible 30,000 quarters of historical noncompliance in the database. EPA accepted the corrections on approximately 2,400 of these compliance quarters. After summarizing this data, the report concluded:

> EPA headquarters and Regional staff consider SFIP data to be of high quality, pointing out that facilities commented on only a "small amount" of the data they reviewed.[7]

In actuality, the volume of requests for corrections was quite high, considering the fact that not all companies track EPA's databases with great care. More importantly, the rate of error in EPA's data, which ran in the 8-11% range by the Agency's own admission, is quite high for a program that is attempting to tell citizens what is occurring in their neighborhoods. By analogy, local public services, such as trash pickup or mail delivery, would be seen as inadequate if they had a failure rate of 10%. Certainly a company marketing a consumer product that exploded one out of ten times it was used could not stay in business for any length of time.

It is at least possible that the error rate found in the SFIP could be viewed as acceptable if the purpose of the project had been limited to assessing broad patterns of environmental compliance. EPA has, however, indicated that the SFIP is intended to provide public access to "facility-level information" and to "improve multimedia facility profiling."[8] When the goal of an information product is to characterize facility-specific circumstances, users of the data are probably expecting much higher levels of accuracy.

3. *Attempting to Serve Multiple Audiences* - Data quality debates also arise when data intended for one audience is offered to another audience that has a different set of questions and needs. This situation arises most commonly with data assembled for EPA's regulatory programs. EPA and the states assemble data about the operations of facilities and collect ambient data from a relatively dispersed set of environmental monitors. The primary audiences for these data are the state regulatory agencies and the EPA Regional offices. The professional staff using these data understand their origins and appreciate the limitations inherent in the current systems for collecting these data. For example, the monitoring network for ambient data is typically not extensive enough to characterize conditions in very local settings.

When these data are posted on the Internet, however, they are available to a much wider group of people who are trying to answer a variety of questions. These potential users of EPA's Web site will not see the data through the eyes of the government program managers, understanding the limitations inherent in the data. These potential users also do not typically have the time or inclination to learn the details of these environmental programs in order to create the proper context for the data's use. These users may also not be interested in the question that the data was intended to answer. As an example, they may be trying to determine

---

[7]Evaluation, at 36.

[8]http://es.epa.gov/oeca/sfi/overview.htm

whether the emissions from a plant are safe, while EPA or a state may be measuring the efficiency of a waste treatment unit, with little concern about the actual safety of an emission.

EPA's Index of Watershed Indicators (IWI) provides a good example of how miscommunication can occur when data intended for measuring regulatory compliance is repackaged and presented to the general public. The IWI has been a component of EPA's "Surf Your Watershed" Website for several years. Under the IWI, watersheds are assigned numeric scores reflecting a rating of the watershed's condition and vulnerability. Of the 18 indicators used in the scoring system, the indicator that measures consistency with state water quality standards receives the greatest weight. The data for this indicator are drawn from the reports that states file under Section 305(b) of the Clean Water Act.[9]

As EPA and state agencies have recognized, the Section 305(b) data are not measuring water quality against a common set of health or environmental benchmarks. Instead the states set water quality standards for different water bodies and then monitor ambient conditions against the relevant standard. These standards will vary from place to place, depending on the state's water quality strategy. In some cases, states will apply stringent water quality standards to relatively pristine waters in order to maintain sensitive ecosystems. In contrast, where the expected use of a waterbody is for industrial use, the state may set much less stringent standards. Given this framework, the IWI can actually give poor ratings to watersheds that have very clean water because those watersheds are not in complete compliance with the stringent water quality standards that are applicable.

The federal and state regulatory agencies that use the Section 305(b) data understand that these data are useful for measuring program management progress, but are not designed to be general "quality of life" measures in communities across the country. This subtlety is easily lost on the average citizen who might try to use the IWI to assess the safety of his local watershed and compare it to other areas. This is particularly the case when EPA suggests to the users that the IWI should be used to measure a community's health. For example the opening page of the IWI on EPA's Web site begins with the following statement:

> The Index of Watershed Indicators (the IWI or Index) is a compilation of information on the "health" of aquatic resources in the United States. Just as a physician might take your temperature & your blood pressure, check your pulse, listen to your heart beat and respiration, evaluate your weight compared to your height, etc., the Index looks at a variety of indicators that point to whether rivers, lakes, streams, wetlands and coastal areas are "well" or "ailing" and whether

---

[9]The data collected for Section 305(b) also suffer from the absence of a standard set of protocols for collecting data. States may have differing methodologies for measuring compliance with their water quality standards. Thus the data are often not comparable between states.

activities on the surrounding lands that affect our waters are placing them at risk.[10]

This quotation clearly presents the IWI as a guide for considering community health issues. There is no suggestion that the IWI's primary indicator is a means to measure the effectiveness of state water quality programs rather than a measure of actual water quality.

The IWI, and other EPA Web sites such as the National Air Toxics Assessment or the Risk Screening Environmental Indicators Model, reflect a systematic effort by the Agency to transform databases holding regulatory monitoring data into resources for the general public about the state of the environment in communities. Such information resources reflect a diffusion of the "TRI ethic" into other EPA programs. In its zeal to pursue these goals, EPA is expecting the general public to have a highly sophisticated appreciation of the distinction between program management measures and indicators of environmental health. The situation has become particularly problematic where, as in the case of the IWI, the Agency has not provided an effective explanation of this distinction.

4. *Maintaining Objectivity in Public Discourse* - The data EPA posts on the Internet are routinely used by other parties in the private and public sector to produce information resources that these parties then present to the public. These derivative resources reflect the views of the third parties on the meaning of EPA's data. In most cases these resources stay within the bounds of reasonable opinion on environmental matters. At times, however, the characterizations of environmental conditions by third parties reflect strong biases that can be misleading.

It is not clear what criteria guide EPA in deciding whether to challenge publicly a third party characterization of the Agency's data. Recent statements by EPA suggest that the Agency policy is to refrain from criticizing third party characterizations of EPA data. In a recent letter from the House Commerce Committee, EPA was asked why it had not questioned the use of the Risk Screening Environmental Indicators by Environmental Defense to calculate facility-specific risk when the Science Advisory Board and the Agency had acknowledged that the Indicators were not appropriate for site-specific risk assessment.[11] In response, EPA offered a remarkable statement:

---

[10]http://www.epa.gov/iwi

[11]Letter to Carol Browner, Administrator of EPA, from Congressman Michael G. Oxley and Congressman Michael Bilirakis (May 4, 2000).

As a federal government agency, EPA neither publicly endorses nor criticizes the analyses of our data and development of mathematical models, or subsequent reporting of our data and adaptation of EPA models by other organizations.[12]

While this statement appears to articulate a broad policy, EPA certainly has expressed its disagreements with the views of other parties on environmental issues, including those involving interpretations of EPA data. In the regulatory context, for example, EPA routinely challenges, usually as part of its response to comments, the conclusions drawn from available Agency data by private parties. Thus CEEI assumes that the Agency does not really intend to abstain from all public comment on third party characterizations of Agency data.

What is significant in the above statement is that EPA apparently believes that its obligation to be neutral and objective implies a higher level of deference to the views of others in its information dissemination activities than in more traditional Agency programs. Many stakeholders, however, believe that objectivity actually requires a more activist Agency role in public discourse about environmental issues. In this view, EPA certainly has an obligation to speak out when it believes that third parties have rendered false or misleading presentations of Agency information. Statements like the one cited above indicate that EPA remains unclear, even ambivalent, on what its role should be in "clarifying the record" in public debates over environmental issues.

5. *Improving vs. Describing Data Quality* - During the last several years, EPA has said that it is committed to "improving" the quality of the information it provides to the public. Some stakeholders have assumed that this commitment includes an active role in obtaining the information necessary to address public needs. For example, when EPA acknowledges that an existing database, which may contain regulatory information about facility compliance, is not well-suited to describe local environmental conditions, will EPA collect additional data to develop statistically valid answers to questions about local conditions?

As a general matter, EPA has not been inclined to collect additional data in new areas that are not directly related to its regulatory mandates. Instead, EPA has decided to post the data it has and provide explanations, often described as "metadata", about how the data was collected. EPA's operating assumption is that the users of the data will understand the limitations of the data and adjust their expectations accordingly.

This somewhat passive approach to "improving" data quality has been unpopular with a variety of stakeholder groups. NGO groups have criticized EPA for the failure to address data gaps. Statistician organizations have suggested that EPA's public databases are not providing

---

[12]Letter to Congressman Michael G. Oxley and Congressman Michael Bilirakis from Diane E. Thompson, Associate Administrator, Office of Congressional and Intergovernmental Relations, EPA, (June 13, 2000), Response to Question 30.

useful information to the public because the databases are not designed with the statistical rigor needed to provide reliable answers to public questions.

Business groups have also been concerned because EPA continues to promote the use of these databases for purposes for which they were not designed, while not undertaking any steps to upgrade the data. The mere attachment of metadata explanations, which can be difficult to understand, to the databases has not really improved the quality of the information offered to the public.

As an example, EPA had expected an ambitious role for information in its implementation of the Government Performance and Results Act (GPRA). Under the GPRA, all federal agencies are required to develop strategic plans that articulated specific goals for their programs. Agencies are expected to develop "outcome-based" measures of program goals to provide a specific metric to judge program progress. Finally, agencies are required to provide public reports annually, beginning in March of 2000, on the progress they are making in achieving their goals.

Initially EPA expressed a strong desire to use the GPRA mandate as a tool to manage its programs better and to provide a clearer public articulation of what its programs were achieving. To that end, EPA was one of the first agencies to redesign its budget into a format organized around the GPRA goals. In addition, EPA established a staff within the Office of the Chief Financial Officer that was to provide leadership on the measures and data that arose around GPRA implementation.

These efforts, however, did not ultimately lead to any new initiatives to assemble additional data. In March of 2000 EPA issued its public report on the GPRA.[13] EPA relied on existing data, primarily regulatory compliance data, that are currently in EPA and state files. The report openly acknowledges the data quality concerns it faces:

> EPA gathers much of its data on the environment from sources outside the Agency, whose reporting cycles and data standards vary widely. For both its own data and those provided by outside sources, EPA must continue to focus on the quality and availability of the data in order to ensure accurate measurement of program results. In a few instances, data relevant to FY 1999 APGs are either lacking or of poor quality.[14]

The GPRA report indicates that EPA plans to work with other "program partners" to improve its measures and the data used for measurement. Yet this general commitment has not been translated into concrete steps to change existing measures or to assemble different data.

---

[13]U.S. EPA, "U.S. Environmental Protection Agency Fiscal Year 1999 Annual Performance Report," EPA-190-R-00-001 (March 2000).

[14] Id., at 7.

Conclusions: The above discussion of the root causes of controversy suggest the following lessons for EPA's future activities on data quality:

a. EPA should make feasible commitments, and then follow through on those commitments. Ambitious goals with weak results only erode the Agency's credibility.

b. Data is seldom "high quality" for all purposes. The meaning of data quality is derived from how data will be used.

c. EPA should be clear about the purposes and appropriate uses of particular data. The Agency does not serve the public by silence on this issue or by reciting vague purposes for a database.

d. Clarifying the audience for information is important and closely linked to the need to define the purpose of an information product.

e. EPA should be willing to challenge third party uses of the Agency's data that represent misuse of the data. The willingness to raise concerns when warranted can enhance EPA's reputation for objectivity.

f. EPA needs to reconcile its data quality and public access objectives. "Government in the Sunshine" policies favor prompt disclosure. Maintaining data quality may require taking the time to "do it right the first time". EPA should distinguish the various types of disclosures it makes (e.g., access to original documents vs. creation of new interpretive resources) when establishing policies that balance these interests.

V. Framework for Data Quality

As documented in the previous sections of this paper, a series of factors have contributed to misunderstandings and controversy about data quality. These controversies exist despite the fact that data quality, as a general goal, is highly valued by most participants in environmental information issues.

CEEI believes that much progress could be made on data quality if stakeholders shared a common framework for addressing data quality goals. Such a framework could facilitate communication on the issues and help to minimize unnecessary conflicts. Accordingly, CEEI provides the following recommendations for a management framework on data quality:

A. Types of Dissemination Activities

The information dissemination activities at EPA tend to fall into one of three general categories that carry differing data quality responsibilities:[15]

1. Management *of original documents* - In the normal course of its business, EPA generates and receives a variety of documents that contain environmental information. EPA must manage these official records of the Agency and meet legal requirements about security and public access.

2. *Compilation of data from other documentary materials* - This type of resource, usually an interactive database, assembles in one place data that is contained in disparate filings, reports, monitoring logs or other documents. The principal value of this type of resource is that it provides a much more efficient means for information users, both inside and outside government, to conduct analyses and research with these data.

3. Development *of interpretive information products* - This activity is intended to create a useful information product that will help an audience understand a situation or make relevant decisions. This type of resource, which is designed to answer particular questions, is the most ambitious form of information dissemination activity.

B. Data Quality Objectives

The different forms of information dissemination activity carry with them different data quality objectives. These objectives need to be clearly defined in each context, and the Agency should develop policies and procedures that will achieve those objectives. The principal data objectives for each activity are as follows:

1. *Management of original documents* -
   - Timely disclosure
   - Responsive (i.e., complete) disclosure
   - Source and origin of documents[16]

2. *Compilation of data from other documents* -
   - Data accuracy (i.e., does the database track the original source?)[17]

---

[15]All three categories of activities generate documents or other pieces of information that may constitute "records" for purposes of laws like FOIA. In accordance with such laws EPA is required to provide for public access upon request to such documents, except to the extent that these documents (or parts of the documents) fall within statutory exemptions.

[16]Where documents do not "speak for themselves", information about where the document originated, its date of creation and other related contextual information can be important to the users.

[17]The accuracy level needs to be relatively high because the potential uses of the data are unpredictable at the time the data are assembled.

- Description of data parameters (e.g., age, scope, data collection methodology)
- General definition of appropriate use[18]

3. *Formulation of an information product* -
   - Application of the Stewardship Principles (see below)

C. Stewardship Principles

When EPA decides to develop an information product that explains environmental circumstances, recommends particular behavior or generally assists decisionmaking by others, the Agency should accept stewardship responsibilities that derive from general product stewardship principles. These responsibilities are also consistent with best practices in risk communication.[19] Effective application of these principles will lead to information products that achieve appropriate levels of data quality.

The essence of the stewardship responsibility is **providing information that is accurate about an important environmental question in a form that is understandable to the customer.** There are at least three key components of this definition. First, the information must be an accurate representation of the situation. Second, the information should be designed to answer questions that are important for public health and environmental protection. This obligation often requires an exercise of judgment about the information that is truly relevant to an environmental question, based on the best available science. Third, the information must be presented in a way that allows the customer to understand and appreciate the significance and context for the information.

A core concept guiding the Information Stewardship Principles is the recognition that the meaning of "data quality" is contextual. There are no absolute terms that can define high quality data for all situations. The adequacy of data quality is dependent on the purpose and audience for the data. Databases that are fully acceptable for one purpose could be inadequate in other

---

[18]While EPA cannot predict every possible use of the data, the Agency should generally describe the types of purposes for which the data is most and least reliable. For example, EPA should clarify the geographic scale (e.g., national, regional, site-specific) for which the data should be used.

[19]For an overview of the principles of risk communication and a summary of existing manuals on risk communication practices, see OECD, "Risk Communication for Chemical Risk Management: an OECD Background Paper" (September 2000).

contexts. The adequacy of data will also be determined by the needs of a particular audience, which may require different degrees of rigor.[20]

Thus the starting point for a sound information product is an articulation of the intended purpose and audience for the product. In defining the purpose, EPA should have a clear picture of the questions to be answered and the possible decisions that the users might make with the information. To appreciate the audience's needs, EPA should understand how the audience uses data and the level of refinement that they are expecting.

Once the purpose and audience parameters are defined, EPA can then establish the expectations of the information product along several dimensions of data quality:

a. *Accuracy* - Recognizing that most data collections are not perfectly accurate, what error level can be tolerated?

b. *Timeliness* - What is the relevant time frame for the audience's use or decision? How close to "real time" should the information be?

c. *Comparability* - What comparisons would the audience like to make? Must the data be consistent across geography? Must the data be consistent over time?

d. *Precision* - How refined is the analysis? What are the implications for sampling or analytical methods? What type of statistical statement does the audience seek? What are the implications for the amount and distribution of data that must be obtained?

e. *Contextual information* - How complex is the question from the audience? What degree of scientific or technical background is necessary to understand the data? What operational, historical or program information is needed to understand the data?

f. *Data required* - What data should be available to meet the information need? Can existing data answer the question? Have existing data been collected under conditions that suggest they will be reliable? Is more information needed?

g. *Process considerations* - Is peer review appropriate? Is public comment appropriate? Should the information product be "beta-tested" with the audience to assure it is understood?

The stewardship obligation includes an ongoing commitment to examine the product's purpose and audience, and the related parameters of data quality, throughout the "life cycle" of the product. These principles guide the initial data collection strategies for the product and the

---

[20]This conclusion is naturally consistent with the Total Quality Management (TQM) definition of quality, which emphasizes meeting customer requirements, rather than some absolute standard.

design of a new product. These principles are also particularly important when EPA tries to use data that was collected for one purpose for a new "secondary" use. Stewardship does not imply that EPA should never recast data for another purpose. Instead, by application of the thought process embodied in the Stewardship Principles, EPA can determine whether a secondary use of particular data is justified. Finally, these principles provide useful criteria for evaluating existing products and will help EPA determine whether changed conditions warrant modification of the product.

A key consideration in the operation of the Principles is the concept of "labeling" for the information product. For most consumer products, labeling is a primary tool by which the product seller identifies the appropriate application of the product, explains how to use the product and offers any needed caveats or descriptions of the product's limits. There is a need for similar communication with customers for information products as well. During the design and evaluation of an information product, it is important for the information steward to determine whether the product's labeling is effective in communicating the right messages to the customer.

While application of the Stewardship Principles can be achieved through a variety of mechanisms, there are some natural steps that are implied by the logic of the Principles. Any information product goes through a design stage in which the developer decides whether the available information can support a viable product that will satisfy the intended purpose and audience. If it appears that a useful product is possible, the developer should create a prototype of the resource and "beta test" it with a sampling of the intended audience to determine whether the purpose is being served. In some cases, it is prudent to test various versions of the product.

Based on the "beta test" results, the developer should make appropriate modifications and develop a final product. In issuing a final product, the developer should also arrange for ongoing maintenance of the product to answer public questions about the product and to make corrections to the data where needed. On a periodic basis the whole product should be evaluated to determine whether it continues to provide a valuable service.

D. The Challenge of Classification

While this framework provides a useful paradigm for defining EPA's data quality obligations, there may be situations where it is difficult to classify a particular information dissemination activity. For example, in some cases it may be challenging to determine whether EPA is assembling a database to assist research or presenting an information product to convey a particular message to an audience. In those cases, it is prudent for EPA to apply a functional test (i.e., how does the public understand and use the data), rather than the form of the resource, when deciding how to classify a particular resource.

The TRI program provides a useful example to demonstrate this challenge because the program has generated a very robust set of information resources for public use. EPA offers the public all three types of information resources described in this framework. First, the original Form R documents for any facility in the TRI system can be obtained from EPA. Second, the TRI data is assembled into a database that is available for the public to view and to manipulate

for analytic purposes.[21] Third, EPA issues at least one national report each year explaining the significance of the TRI data and has developed interactive resources, such as the Risk Screening Environmental Indicators, that link TRI data to other information on chemical hazard and exposure. In addition, NGO groups have developed a wide range of interpretive reports and Web resources based on the TRI data.

Supporters and critics of the quality of the TRI data have not always been clear about which dimensions of data quality they are concerned about. This has led to miscommunications between these groups. For example, supporters of the TRI data focus on the extensive quality assurance steps EPA takes to assure that the data contained in the public TRI database accurately reflects the Form R submissions. These quality control steps have, in fact, assured that TRI has a very high level of accuracy relative to other EPA databases. Yet there are other dimensions of data quality on which TRI is less successful. For example, those who have criticized the popularity of TRI data emphasize that the data is often two years old by the time EPA provides the public release of the data.[22]

While there has been substantial debate about the significance of TRI data in the public mind, there is only limited research on how the TRI data actually reach the public and how the public interprets their significance. Better insights into that question would be quite helpful in deciding when and how the Stewardship Principles should be applied in the TRI program.

7099616

---

[21]Increasingly Envirofacts has become the common venue for users of the TRI data.

[22]Most of this delay is inherent in the statutory design of the TRI program. Submitters compile the data for an entire calendar year on the Form R's and then submit those forms six months into the following year. Thus TRI data can be as much as 18 months old on the day it is submitted. The processing of the myriad Form R's takes EPA an additional six to nine months.

Before the United States
Office of Management and Budget

Proposed Guidelines for Ensuring and Maximizing the Quality, Objectivity, Utility and Integrity of Information Disseminated by Federal Agencies

---

Comments of the Coalition for Effective Environmental Information

---

On June 28, 2001, the U.S. Office of Management and Budget (OMB) issued a Federal Register notice (66 Fed. Reg. 34489) requesting comment on proposed Guidelines for Ensuring and Maximizing the Quality, Objectivity, Utility and Integrity of Information Disseminated by Federal Agencies. These guidelines, which are intended to implement Section 515 of the Treasury and General Government Appropriations Act for Fiscal Year 2001 (Pub. L. No. 106-554), will be incorporated into OMB's Circular A-130, "Management of Federal Information Resources". OMB intends that agencies will make appropriate changes to their regulations and policies to implement the provisions in these Guidelines.

The Coalition for Effective Environmental Information (CEEI) is a group of major companies and business organizations, representing a wide range of industrial sectors, that share a common interest in improving how the government collects, manages, uses and disseminates environmental information.[1] CEEI supports public policies that encourage data quality, governmental accountability, efficient data collection, alignment of data with strategic goals and consistent management of environmental information resources.

The proposed OMB Guidelines address one of CEEI's primary interests – improving the quality of information that the federal government offers to the public. We support efforts by OMB and individual agencies to address data quality issues in a systematic way. By way of background, we attach a document, submitted to the Bush Administration in December of 2000, that presents a series of recommendations for improving environmental information management. Many of these recommendations include good practices that have applicability beyond the environmental field.

The attached document gives special attention to the topic of data quality. In addition to providing specific recommendations for improving data quality, the document includes an appendix titled "The Data Quality Challenge" which provides our analysis of some root causes of disputes about data quality and a framework for how these challenges can be addressed by federal agencies. We believe that the perspectives reflected in our analysis have equal relevance to information policies in other federal agencies. We strongly urge OMB to consider this analysis as it proceeds with these Guidelines.

---

[1] CEEI includes representatives from the aerospace, chemical, energy, automobile, pharmaceutical, forest products, petroleum, electronics and consumer products industries.

Based on its experience with information policy in the environmental area, CEEI offers the following comments on the OMB Guidelines:

Data Quality is a Broad Concept

OMB should maintain a broad definition of data quality. Our experience has taught us that different stakeholder groups often have different perspectives on what are the most important dimensions of data quality. (See The Data Quality Challenge at 5-7.) Some groups focus on the pure accuracy or statistical rigor of data. Other groups emphasize the timeliness or completeness of data. Federal policies should be designed to foster all of these dimensions of data quality. No useful purpose is served by crafting a narrow definition of data quality. At the same time, if OMB decides to embrace explicitly a broader definition of data quality, terms like "completeness" and "timeliness" would need greater definition.

The Importance of Context

It is equally important for the Guidelines to emphasize that agencies should place publicly disseminated data in context with other relevant information. This principle is noted in paragraph V.1.B of the Guidelines. In the absence of contextual information, the public can often draw unwarranted conclusions.

As an example from the environmental field, distortions can occur when government agencies publish lists of chemicals that they intend to review. In some cases, parties reach conclusions about the chemicals before the review is complete. This has occurred with the federal efforts to assess chemicals, using validated tests, for potential endocrine disruption effects.[2] The U.S. Environmental Protection Agency (EPA) is still in the process of developing validated tests to identify such chemicals. Even though such tests have not yet been established, some organizations (e.g., Illinois EPA – but not U.S. EPA) have already published lists of chemicals that assertedly do, in fact, disrupt endocrine systems in the body. These lists do not make clear that it is not technically possible, at this stage, to identify endocrine-disrupting chemicals with any confidence.

This type of misunderstanding can be mitigated when agencies provide better contextual information. EPA has, for example, provided appropriate contextual information about a list of chemicals under review in the Federal Register notice for its Voluntary Children's Chemical Evaluation Program (65 Fed. Reg. 81700, December 26, 2000). Another positive example is EPA's Risk Screening Environmental Indicators (RSEI) model, which has a page on "strengths and limitations." See http://www.epa.gov/oppt/env_ind/strength.htm.

---

[2] This review was mandated under the Food Quality Protection Act of 1996.

Three Types of Information Activities and Corresponding Quality Obligations

In paragraph III.1 of the Guidelines, OMB states that "Quality is to be ensured and established at levels appropriate to the nature of the information to be disseminated." This is an extremely important principle to establish, and OMB should consider amplifying the concept. As described in more detail in the attached document, CEEI has found it useful to distinguish three types of agency activities that carry differing levels of responsibility for data quality: (1) management of original documents; (2) compilation of data from other documentary materials; and (3) development of interpretive information products. See The Data Quality Challenge at 15-16.

When an agency is managing original documents submitted by outside parties or generated within the agency, data quality responsibilities are at a minimum. The documents speak for themselves. Agencies are generally responsible for managing those documents in accord with federal policies about security (e.g., Trade Secrets Act), public disclosure (e.g., Freedom of Information Act) and record retention.

Agencies also undertake efforts to consolidate data from separate documents into consolidated databases for use by government employees and, increasingly, by members of the public. Data quality responsibilities become more significant for such resources. The agency must assure that data is accurately transferred to the consolidated data base, that the limits of the original data collection activity are explained and that the appropriate uses of the data are understood.

Increasingly, agencies are also engaged in the development of more sophisticated information resources, often in the form of Web sites, that consolidate different types of information to convey a message or interpretation about a public concern. These "information products" can be quite influential in affecting private sector decisions about companies, facilities and products, particularly when they take the form of Web sites. (As OMB notes, the Internet "increases the potential harm that can result from dissemination of information that does not meet ... quality standards." 66 Fed. Reg. 34490.) The potential influence of these information products in the global marketplace warrants a close look at data quality issues before the agency issues such a product. We believe that agencies developing or significantly modifying interpretive information products should apply "stewardship principles" that are described below.

Information Stewardship

In the Discussion section following paragraph III.2., OMB states that "Agencies may want to consider developing different processes to address different types of information."[3] CEEI agrees with this conclusion and suggests that it be amplified. When an agency is

---

[3] Similarly the existing Circular A-130 states that agencies "will [c]onsider the effect of their activities on members of the public and ensure consultation with the public as appropriate." See §8.a.1.b.

developing an interpretive information product, as outlined above, CEEI believes that "stewardship principles" should guide behavior.

The objective of these principles is to provide accurate information about an important question in a form that is understandable to the customer. The starting point for achieving this goal is to define the purpose and audience for the product. Based on that understanding, the agency can determine what is acceptable under the differing dimensions of data quality (i.e., accuracy, timeliness, precision, comparability, context, data requirements and process.)

In designing an information product, agencies can then define both the content of the information product as well as the "labeling" for the product, which articulates the product's purpose, directions for its use and appropriate caveats. The agency should consider sharing the product with its intended audience and data source provider (i.e., "beta-testing") to see whether the intended messages and applications are correctly interpreted and understood. Agencies should also provide for ongoing maintenance of an information product once it is issued, including such functions as error correction and technical assistance.

It should be noted that implementation of these principles does not necessarily require elaborate procedural steps. What is more important is that agencies go through the thought process required for good information stewardship, which can occur through a variety of mechanisms. A good example of how EPA has begun to institutionalize that thought process is the Agency's "Lessons Learned about Designing, Developing and Disseminating Environmental Information Products" (November 17, 2000). See http://www.epa.gov/oeiinter/pdf/OIAA_Lessons-learned.pdf.

Finally, CEEI supports OMB's conclusion that the quality of an information product should be reviewed and documented before the product is disseminated. (Paragraph III.2.) EPA's Lessons Learned report reached a similar conclusion that data quality steps should begin at the inception of information product development. Id. at 9-15.

Agency Administrative Mechanisms Should Address Conformance with *Agency* Guidelines

In paragraph III.3. and Section IV. of the Guidelines, OMB proposes a series of procedural requirements for agencies to follow. These include mechanisms for citizens to seek correction of information and for agency reporting to OMB on the resolution of data quality complaints. Under the proposed Guideline language, these mechanisms and reports to OMB address complaints about whether the agencies have complied with the OMB Guidelines.

Yet we understand that, under Public Law No. 106-554 and Circular A-130, each agency would create its own data quality guidelines incorporating the OMB Guidelines. It would seem, then, that the touchstone for the agency correction mechanisms and reporting obligations would be the *agency policies and regulations* that are established to implement the general policies of the Guidelines. Accordingly, we suggest that any information correction mechanisms or reporting obligations be tied to the data quality standards in the agency policies and regulations, rather than the OMB Guidelines themselves, which are not intended to be self-implementing.

We also recommend that OMB define some reasonable time frame (e.g., 60 days) within which a correction request should be addressed.

Agencies Should Seek Public Comment on their Draft Guidelines

The Guidelines do not clarify whether agencies are required to seek public comment on the policies and regulations that they establish to implement these Guidelines. A wide range of stakeholders will very likely be interested in the data quality policies of individual agencies – potentially even more than they are in OMB's guidelines. Accordingly, CEEI recommends that the OMB explicitly call for public comment on the regulations and policies that agencies create to implement the Guidelines.

At a broader level, OMB should be requiring greater transparency in all agency processes intended to implement these Guidelines. It is important for the public to know how they can participate in the development and implementation of data quality policies in the agencies. It is also important for OMB to make clear that agencies have a responsibility to respond to public inquiries and comments about these issues and to explain final decisions they make in response to correction requests.

Approach to Hyperlinks

The definition of "information" in the Guidelines excludes "hyperlinks to information others disseminate." We have assumed that this language is intended to avoid a suggestion that federal agencies are required to conduct data quality reviews of all sites to which they might link their Websites. CEEI recognizes the legitimacy of that concern and does not favor policies that discourage linking public and private sector sites. Creation of information networks is a valuable aspect of the Web environment.

At the same, federal agencies can sometimes be faced with complex policy questions about whether they should link their Web sites to sites managed by other parties. Data quality could be one of many factors that they should consider in making those decisions. Federal agency policy on the appropriateness of hyperlinking to other Web material is not well-defined, and there has been little public discussion about what policies make sense. We cannot see that an agency decision about hyperlinking to another site would never be germane to the data quality policies of the federal government. Thus we would recommend deleting this exclusion for now, pending a maturation of federal policies in this area.

Breadth of the Data Quality Obligations

The OMB Guidelines are understandably focused on the quality of information that federal agencies intentionally develop and disseminate to the public. As noted in the definition of "dissemination" in Section V., the Guidelines are not applicable to "responses to requests for agency records under the Freedom of Information Act." This limitation on the scope of the Guidelines is a reasonable one that CEEI does not challenge.

In looking at the longer term implementation of these Guidelines, however, the federal government should consider how data quality responsibilities will be aligned with the move toward Web postings of government information that results from the Electronic Freedom of Information Act (E-FOIA). Federal law now requires that government records that "have become or are likely to become the subject of subsequent requests for substantially the same records" become "reading room records", requiring agencies to make them available "by computer telecommunications or . . . other electronic means."[4] This mandate is pushing agencies toward Web posting of their internal documents and databases. As the federal government moves steadily forward toward an "e-government" mode of operation, the distinction between internal and public documents will continue to blur.

What are the implications of this phenomenon for improving the quality of information that the public receives from federal agencies? CEEI believes that one implication of this shift is that agencies should institutionalize stewardship principles wherever possible, leading to better quality information in all aspects of federal operations. Today's internal database for a limited purpose may become tomorrow's major public information resource.

At this time, however, CEEI does not recommend any explicit changes in the Guidelines to anticipate these changes. Instead, we urge OMB to monitor the implementation of E-FOIA, and the move toward "e-government" in general, to determine what types of information the agencies are moving to the Web and the respective quality of that information. Explicit policies to address these trends may make sense some time in the future.

Conclusion

CEEI appreciates this opportunity to comment on OMB's Guidelines. We look forward to timely issuance of the Guidelines and full implementation of the Guidelines at federal agencies. We believe that a more systematic approach to data quality at federal agencies provides a valuable service to the public and enhances the credibility of public institutions.

Coalition for Effective Environmental Information
August 7, 2001

---

[4] 5 U.S.C. §552(a)(2)(D).